

Matière Première
Revue d'épistémologie
N° 1/2010

Épistémologie de la médecine et de la santé



Sous la direction de

Gérard **Lambert**
&
Marc **Silberstein**

Nicolas Baumard / Cyrille Bouvet / Joël Coste / Catherine Dekeuwer
Steeves Demazeux / Michael Esfeld / Christian Frelin / Élodie Giroux
James A. Marcum / Jérôme Ravat / Bernard Swynghedauw

Collection
Sciences
& Philosophie
Éditions Matériologiques
materiologiques.com

Ebook offert par les Editions Matériologiques, 23 mars 2020

Matière Première

Revue d'épistémologie

N° 1/2010 (nouvelle série)

Épistémologie de la médecine et de la santé

Éditions Matériologiques
www.materiologiques.com

[Comité de rédaction de la revue Matière Première]

[**PASCAL CHARBONNAT**] professeur de philosophie, docteur en philosophie des sciences, Paris X / [**JEAN-MARC DEL PERCIO**] historien, Saint-Jeure d'Andorre / [**GÉRARD LAMBERT**] médecin, écrivain, journaliste médical, Paris / [**GUILLAUME LECOINTRE**] phylogénéticien, professeur au Muséum national d'histoire naturelle, Paris / [**FRANÇOIS PÉPIN**] professeur de philosophie, docteur en philosophie des sciences, Paris X / [**VALÉRY RASPLUS**] sociologue / [**MARC SILBERSTEIN**] éditeur, fondateur de *Matière première*.

[Comité scientifique de la revue Matière Première : voir le site des EM]

[AVERTISSEMENTS AUX LECTEURS]

Ce PDF contient des liens hypertextes, permettant d'accéder par un simple clic à des sites internet ; ainsi, le lecteur aura à sa disposition de nombreux compléments tels que des notices biographiques, des textes sources, des articles connexes, de l'iconographie, etc.

Ces liens sont signalés par un cadre rouge. Il suffit de cliquer dessus et d'attendre l'ouverture du navigateur.

Pour joindre la revue (propositions de textes, de thèmes, renseignements divers, etc.), merci d'utiliser exclusivement notre adresse électronique :
contact@materiologiques.com

Matière Première. Revue d'épistémologie (n° 1/2010)

ISSN: 2110-5650

© **Éditions Matièreologiques**, octobre 2010.

c/o François Pépin, 233, rue de Crimée, F-75019 Paris

www.materiologiques.com

contact@materiologiques.com

Conception graphique, maquette, PAO, composition, photos de couverture et de l'intérieur :

© Marc Silberstein.

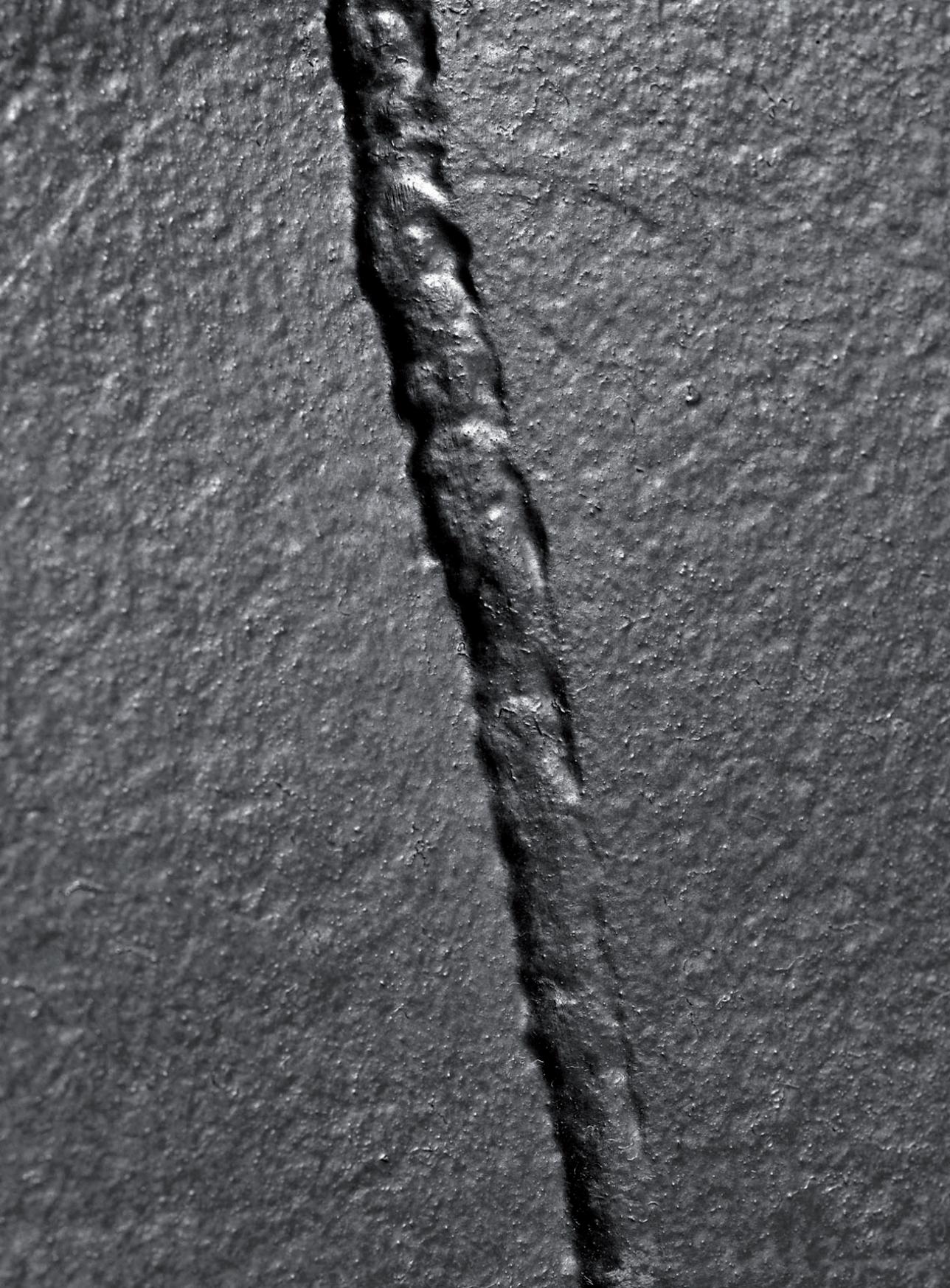
Table des matières

Thème [Médecine et santé]

- page 5 **Introduction** [Gérard Lambert & Marc Silberstein]
- page 9 **Facteurs de risque et causalité en épidémiologie**
[Élodie Giroux]
- page 31 **Le concept de fonction dans le discours
psychiatrique contemporain** [Steeves Demazeux]
- page 75 **Causalité(s) en psychopathologie : des modèles
inconciliables ?** [Cyrille Bouvet]
- page 101 **Définir la maladie génétique** [Catherine Dekeuwer]
- page 129 **Cancer : complexity, causation, and systems biology**
[James A. Marcum]
- page 153 **L'évolution biologique en médecine. Un outil
de formation, un moyen de hiérarchiser les informations
et une base rationnelle à une politique de santé** [Bernard
Swynghedauw & Christian Frelin]
- page 177 **Inférences causales et probabilistes dans le
raisonnement diagnostique : un éclairage historique sur
le débat contemporain** [Joël Coste]

[Varia]

- page 199 **Les fondements de la causalité** [Michael Esfeld]
- page 223 **Fonction naturelle et fonction morale : similitude
ou imposture ?** [Jérôme Ravat]
- page 243 **Une théorie naturaliste des phénomènes moraux
est-elle possible ?** [Nicolas Baumard]
- page 287 **Le matérialisme considéré du point de vue
cosmologique et évolutionniste** [Michel Paty]



Gérard Lambert & Marc Silberstein **Introduction**

Le premier volume de *Matière première*, nouvelle série, est principalement consacré à l'épistémologie de la médecine et de la santé. Sept articles posant des problématiques très différentes permettent d'aborder quelques-unes des questions qui parcourent ce champ de réflexion aux multiples implications scientifiques, philosophiques et éthiques. La causalité, l'étiologie, la fonction sont les notions cruciales questionnées ici.

L'article d'**Élodie Giroux**, « **Facteurs de risque et causalité en épidémiologie** » examine les théories et concepts relatifs à la relation entre une pathologie (P) et un facteur (F), déterminant ou de risque. La question centrale est celle-ci : F est-il *la* cause de P ? Ou encore F est-il un élément d'un faisceau factoriel multiple, dont les f (sous-facteurs) sont, en fonction des circonstances (écologiques, sociales, immunitaires, etc.), convergents, synergiques, en interaction, etc., pour, en dernière analyse, être considérés comme des causes *stricto sensu* ou bien, pourrait-on dire, des quasi-causes ?

Dans « **Le concept de fonction dans le discours psychiatrique contemporain** », **Steeves Demazeux** aborde la double question de la fonction en médecine psychiatrique et celle de la pertinence des catégories nosographiques en psychiatrie. La notion de fonction, éminemment problématique, est centrale en biologie, notamment en biologie de l'évolution. En psychiatrie aussi, comme l'auteur le démontre, elle ne va pas sans poser d'énormes problèmes de définition et de délimitation. Il faut d'abord la débarrasser de ses attributs téléologiques pour en faire un outil d'analyse et de compréhension dûment pertinent. Définir la fonction, c'est donc définir la dysfonction, problématiser l'une, c'est problématiser l'autre. Or, les

notions de trouble psychiatrique, de fonction mentale normale et donc de maladie mentale sont redevables de ces saisie épistémologique d'une réalité souvent mouvante, fluctuante, systémique. L'épineux et permanent problème de la démarcation entre normal et pathologique demeure et si les affirmations déjà anciennes de l'antipsychiatrie et des odes à la renonciation à l'objectivation de la maladie mentale sont bel et bien caduques, la psychiatrie, pour être au plus près de l'impératif éthique du meilleur soin et de la nécessité épistémologique de la meilleure science, semble devoir trouver un chemin permettant d'éviter de tomber de Charybde (de la subjectivité absolue de la maladie mentale) en Scylla (de la déshumanisation de la maladie mentale). L'article produit un examen critique des propositions des principaux acteurs de ce débat, et conclut que «peut-être le mieux est-il de renoncer tout bonnement à la perspective de fixer un critère biologique général et définitif du normal et du pathologique valable pour l'ensemble des troubles mentaux».

Dans «**Causalité(s) en psychopathologie : des modèles inconciliables ?**», **Cyrille Bouvet** interroge lui aussi les notions de maladie mentale, de trouble psychique, les catégorisations psychopathologiques et les théories sous-jacentes. Encore une fois, ce qui est abordé ici, c'est la pertinence d'une ontologie discrète (discontinue) de la maladie *versus* une critériologie établie sur un continuum normal-pathologique. Le parti pris de Cyrille Bouvet est semble-t-il assez minoritaire dans le milieu des praticiens de la santé mentale, versant psychopathologique, et non pas psychiatrie; en effet, il défend l'idée forte et régulatrice que le soin, en ce domaine comme dans les autres secteurs de la médecine (au sens large), ne peut se soustraire – fût-ce pour des raisons de pragmatique thérapeutique – à une tendance à la scientification de ses démarches, méthodes et objets.

Face à l'inflation permanente à de prétendues découvertes de gènes pour à peu près tous les phénomènes biologiques, pathologiques ou normaux, il s'avère utile de rendre compte des réflexions menées par les biologistes ou les philosophes des sciences à propos d'une entité trop souvent considérée comme allant de soi, à savoir le gène. L'article de **Catherine Dekeuwer**, «**Définir la maladie génétique**», est donc consacré à cette question, avec celles, connexes, de la définition de la maladie – nous y revenons toujours – et des gènes comme déterminants de la maladie. En cela, cet article rejoint la problématique traitée dans le texte d'Élodie Giroux. Les maladies génétiques *stricto sensu* sont rares, eu égard

au nombre des pathologiques dans lesquels des facteurs génétiques sont impliqués. Pourtant on appelle de la même façon ces groupes hétérogènes de pathologies. Des mêmes mots étiquettent des réalités empiriques variés, des phénomènes référentiellement distincts.

L'étiologie du cancer est encore largement dominée par la conception dite TMS (théorie de la mutation somatique) dans laquelle des gènes mutés constituent la cause du cancer. Malgré cette prédominance, et les promesses du décryptage du génome humain, de nombreuses anomalies et contre-exemples rendent cette théorie de plus en plus fragile quant à sa portée explicative. Lorsque l'on est favorable à l'idée que la thérapeutique doit être conditionnée par la compréhension des mécanismes des phénomènes en jeu et donc ne plus seulement être le fruit d'une induction plus ou moins longuement menée, au fil de la pratique clinique, ou encore d'une accumulation de faits difficilement compréhensibles dans leur diversité et hétérogénéité – réalisant ainsi pleinement un idéal bernardien –, ces lacunes sont particulièrement frustrantes eu égard aux considérables enjeux de santé public. Aussi, des alternatives théoriques arrivent à maturité, mettant en lumière des circuits causaux non plus linéaires comme dans la TMS, mais faisant intervenir des niveaux d'organisation en interaction. Ces théories relèvent de la biologie des systèmes. Elles sont examinées par **James A. Marcum** dans « **Cancer : complexity, causality, and systems biology** ».

Le cancer, comme bien d'autres pathologies, requièrent le regard conjoint de la biologie fondamentale et de la médecine, des sciences biomédicales. Pour trivial que semble être ce propos, certains domaines de la biologie et la médecine semblent ne toujours pas entretenir le fructueux dialogue que l'on est en droit d'attendre. Dans « **L'évolution biologique en médecine. Un outil de formation, un moyen de hiérarchiser les informations et une base rationnelle à une politique de santé** », **Bernard Swynghedauw & Christian Frelin** regrettent cet état de fait, en prenant le cas de la théorie de l'évolution, trop peu enseignée au cours du cursus médical, et de la médecine darwinienne – quoi qu'on puisse penser de cette dénomination peu satisfaisante – qui est au moins un laboratoire d'idées certainement profitable à la compréhension de certaines classes de phénomènes biomédicaux.

À l'instar de nombreuses autres disciplines biologiques, et suivant une tendance lourde de ces domaines, la médecine doit faire face à une augmentation pléthorique des données qu'elle a à traiter et interpréter.

La multiplication des appareils de mesure des multiples paramètres biologiques qui sont intégrés dans des modèles probabilistes de l'approche de la maladie l'expliquent. **Joël Coste**, dans « **Inférences causales et probabilistes dans le raisonnement diagnostique : un éclairage historique sur le débat contemporain** », prend acte de cette situation mais s'intéresse avant tout à l'histoire des théories et pratiques diagnostiques.

* * * * *

Quatre articles composent la partie «Varia»: un texte de **Michael Esfeld**, « **Les fondements de la causalité** », dans lequel l'auteur établit une épistémologie générale de la notion de cause, afin de parvenir à «trancher le débat philosophique sur les fondements de la causalité en faveur d'une théorie causale des propriétés, voire des structures, comme existant dans le monde».

Les deux autres questionnent la philosophie de la morale et ses rapports avec la biologie de l'évolution, les sciences cognitives, la psychologie: **Jérôme Ravat**, « **Fonction naturelle et fonction morale** » et **Nicolas Baumard**, « **Une théorie naturaliste des phénomènes moraux est-elle possible ?** ». Nous retrouvons, tout comme dans la partie «Médecine et santé», la prégnance des interrogations sur les notions de cause, de fonction, de réduction.

Le dernier, de **Michel Paty**, « **LE MATÉRIALISME CONSIDÉRÉ DU POINT DE VUE COSMOLOGIQUE ET ÉVOLUTIONNISTE** », concerne directement la question des contours et de la portée du matérialisme scientifique.

Élodie Giroux **Facteurs de risque et causalité en épidémiologie**

Fumer est un **facteur de risque** pour un grand nombre de maladies ; l'association avec le cancer du poumon est particulièrement forte. Pourtant, cette association n'est ni nécessaire ni suffisante puisque l'on peut bien être atteint de cette maladie sans avoir jamais fumé et inversement. Or il est désormais acquis que fumer, sans être *la* cause, est toutefois *une* cause du cancer du poumon. Que faut-il dès lors entendre par *cause* et surtout, que comprendre du statut causal de la notion de facteur de risque ? Définie en épidémiologie comme une variable associée statistiquement à la survenue d'une maladie, la notion de *facteur de risque* connaît un usage qui s'est aujourd'hui étendu bien au-delà des frontières de cette discipline. Elle est désormais considérée comme un outil performant pour appréhender la multifactorialité d'un phénomène, dans de multiples domaines d'étude, en particulier les sciences humaines et sociales (PERRETI-WATEL 2004). Quel rôle joue-t-elle dans l'explication causale ? Se substitue-t-elle à la notion de cause ? Ou renvoie-t-elle au contraire à une certaine conception de la causalité ? Au sein même de l'épidémiologie, l'usage de cette notion est ambivalent et imprécis. Elle est parfois utilisée comme synonyme de *risque*, *déterminant* ou *cause*, ou au contraire, bien distinguée, comme simple *marqueur de risque*, de ce qu'il faudrait désigner par *facteur causal* ou *déterminant* (LAST 2004, p. 103).

L'objectif de cet article est d'apporter quelques clarifications sur la relation des notions de *facteur de risque* et de *cause*. Après avoir montré l'origine épidémiologique des facteurs de risque dans le développement des enquêtes étiologiques d'observation – origine qui aurait pu laiss-

ser penser que son adoption entraîne l'abandon du concept de cause au profit de la corrélation statistique –, nous montrerons comment la controverse sur le statut causal du tabagisme dans le cancer du poumon a conduit à la formalisation de critères pour l'inférence causale. Dans un troisième et dernier temps, nous nous interrogerons sur le statut de ces critères qui sont aujourd'hui devenus la boîte à outil de l'analyse étiologique en épidémiologie : constituent-ils avant tout un outil méthodologique efficace qui permet d'éluder la difficile question de la définition de la cause ou cette dernière question demeure-t-elle implicitement présente ?

1 ⇨ Facteurs de risque : modélisation de la multifactorialité et prédiction

1.1 ⇨ Origine et méthode d'identification des facteurs de risque en épidémiologie

L'expression *risk factor* ou *factor of risk* a son origine dans les assurances anglo-saxonnes. Dès les années 1920, des assurances vie aux États-Unis utilisent cette expression à propos d'attributs ou caractéristiques individuelles et physiologiques comme le surpoids et l'hypertension. En épidémiologie, l'expression a commencé d'être utilisée et a été plus particulièrement diffusée et vulgarisée dans le cadre de l'enquête de cohorte d'épidémiologie cardiovasculaire dite de « Framingham » (LAST 2004, p. 103)¹. Quant au concept, celui d'une variable associée statistiquement à la survenue d'une maladie, il est antérieur au XX^e siècle. Dans *Les Causes de la mort* (1989), Anne Fagot-Largeault a bien montré que la recherche étiologique fondée sur l'usage du raisonnement probabiliste a des origines au XIX^e siècle. Toutefois, c'est principalement au milieu du XX^e siècle qu'une notion épidémiologique de *facteur de risque* émerge et s'inscrit au sein d'une recherche étiologique systématisée et encadrée ; puis, cette notion ainsi que l'évaluation des risques en médecine se sont généralisées dans les années 1970 (SKOLBEKKEN 1995). En effet, cette notion, et plus généralement la recherche étiologique en épi-

1. Un des premiers usages médicaux de cette expression daterait d'une publication de l'enquête de Framingham (KANNEL *et al.* 1961). Au sujet de l'importance de cette enquête dans la vulgarisation de « l'approche facteurs de risque », cf. ARONOWITZ (1999), OPPENHEIMER (2006) et GIROUX (2006).

démiologie, ont trouvé un fondement théorique et méthodologique dans le contexte du développement d'enquêtes étiologiques dont la méthode et l'analyse reposent sur des techniques empruntées à la statistique inférentielle². C'est plus particulièrement au lendemain de la Seconde Guerre mondiale aux États-Unis et en Angleterre que des enquêtes de population se sont constituées comme entités méthodologiques à part entière, dans le but précis d'étudier l'étiologie de *maladies dites chroniques* (GIROUX 2006). Dans les pays les plus développés, des maladies qui se caractérisent par une complexité étiologique et une progression lente et progressive, les cancers et les maladies cardiovasculaires en premier lieu, étaient devenues les premières causes de mortalité. Les mécanismes de ces maladies échappaient à l'analyse physiopathologique. Du fait de l'absence de connaissances solides sur leur pathogenèse mais aussi à cause de la nature apparemment irréversible de ces maladies, il apparut d'emblée important à certains médecins de développer des moyens de les dépister au plus tôt et de les prévenir. Par ailleurs, la comparaison des taux de mortalité de différents pays ainsi que les associations statistiques mises en évidence à partir des données des assurances sur la vie conduisirent à formuler des hypothèses étiologiques. La mortalité cardiovasculaire apparaissait notamment plus grande aux États-Unis qu'au Japon : le mode de vie américain et, en particulier, le mode d'alimentation, très différent entre ces deux pays, pouvaient-ils être en cause ?

Mais les associations statistiques des assurances sont biaisées puisque la population à partir de laquelle elles sont établies est sélectionnée. Et par ailleurs, les comparaisons des taux de mortalité se situent au niveau écologique³ et ne permettaient pas d'inférence solide. Pour étu-

-
2. Sous l'impulsion notamment de Ronald Aylmer Fisher, une théorisation de l'inférence dans le cadre d'expériences contrôlées pour les sciences empiriques mettant la statistique mathématique au service de l'expérimentation s'est développée dans les années 1920-1930. La statistique inférentielle constitue alors un outil permettant de contrôler l'observation et ainsi de simuler et même de reproduire les conditions expérimentales de laboratoire (FISCHER 1935).
 3. Le terme «écologique» est utilisé pour désigner le niveau populationnel d'analyse. On distingue généralement les études de comparaison *entre populations* (études écologiques) des études de comparaison *interindividuelle*.

dier rigoureusement ces hypothèses, des enquêtes étiologiques portant sur des individus sont mises en place. Ces enquêtes consistent en une observation organisée dans une population contrôlée. La validité de la comparaison dépend de la neutralisation par le chercheur des variables perturbatrices au moyen d'ajustements statistiques. Elle repose aussi sur la rigueur méthodologique du plan d'étude et, en particulier, du contrôle des biais de sélection et de classement dans la constitution de la population d'étude. On distingue classiquement deux principaux types d'enquête étiologique en fonction des groupes d'individus comparés pour l'examen de l'hypothèse étudiée. Dans une *enquête cas-témoins*, on compare l'exposition passée au facteur étudié dans un groupe de malades et un groupe témoin de personnes qui ne présentent pas cette maladie. Dans une *enquête de cohorte*, on mesure l'exposition à des facteurs à un moment donné du temps et on étudie son devenir du point de vue de la maladie, ce qui permet de comparer les taux d'incidence⁴ dans les groupes d'individus exposés et non exposés (ou ayant connus des niveaux différents d'exposition). Dans l'enquête de cohorte une mesure directe de l'association entre le facteur présumé et son effet est donc possible : c'est le calcul du *risque relatif* qui n'est pas autre chose que le rapport du taux d'incidence dans le groupe exposé à une variable étudiée à celui dans le groupe non exposé⁵. Cette mesure permet d'identifier l'existence d'une relation et donne d'importantes informations sur sa force. Le test de signification statistique permet d'établir si l'association en question n'est pas due au hasard. Ainsi par exemple, à l'issue d'un suivi de vingt années de médecins de sexe masculin, les épidémiologistes britanniques Richard DOLL & Richard PETO (1976) ont pu établir que le risque relatif de décès par maladie cardiovasculaire pour des sujets masculins de moins de 65 ans est : 1 pour les non fumeurs ; 1,7 pour les fumeurs de 1 à 14 cigarettes (ce qui signifie que fumer de 1 à 14 cigarettes multiplie le risque par 1,7) ; 2,2 pour les fumeurs de 15 à 24 cigarettes ; 2,6 pour les fumeurs de 25 cigarettes et plus. La quantification et la possibilité

4. Le taux d'incidence est le nombre de nouveaux cas de maladie sur une période donnée de temps.

5. Dans l'enquête cas-témoins, la mesure de l'association n'est pas *directe* car il n'est pas possible de calculer les taux d'incidence mais il est toutefois possible de calculer un équivalent du risque relatif : c'est l'*odds ratio*. Cf. CORNFIELD (1951).

de pondérer et ainsi de hiérarchiser les facteurs identifiés font la différence par rapport aux dispositifs de la recherche étiologique utilisés avant ce genre d'enquêtes. Outre la possibilité nouvelle de quantifier et hiérarchiser l'association statistique, ce dispositif permet d'appréhender de manière efficace la complexité causale des maladies chroniques et de prédire, en vue de prévenir, leur probabilité de survenue chez des individus.

1.2 ⇨ Les équations multivariées : modélisation de la multifactorialité et prédiction du risque

La pertinence de la notion de «facteur de risque» pour appréhender la multifactorialité s'est plus particulièrement révélée dans le cadre de l'étude de l'étiologie des maladies cardiovasculaires. Si dans le cas du cancer du poumon, une hypothèse étiologique apparut vite l'emporter (le tabac), pour les maladies cardiovasculaires, on a affaire à un grand nombre de petits facteurs jouant un rôle très partiel dans le risque : hypertension, hypercholestérolémie, âge, sexe, tabagisme, présence de diabète, etc. L'analyse des données des grandes enquêtes cardiovasculaires de cohorte rendit vite nécessaire l'usage de modèles mathématiques pour la comparaison et le calcul des risques relatifs des différents facteurs étudiés : en effet, les tableaux de comparaison à double entrées (taux d'incidence dans le groupe d'exposé et dans le groupe non exposé à la variable étudiée) furent rapidement insuffisants pour la comparaison et la quantification (GIROUX 2008). Des modèles d'analyse multivariée sont alors utilisés et adaptés aux données d'observation de l'épidémiologie (CORNFIELD 1962). Les modèles de régression permettent de modéliser un phénomène (variable dépendante Y) par une fonction de plusieurs variables indépendantes (X) : $Y = \alpha + \beta_1 X_1 + \dots + \beta_n X_n$. Il est alors possible de prendre en compte la combinaison d'un grand nombre de facteurs dans le risque et leurs interactions, de contrôler les éventuels biais de confusion et de s'assurer de la contribution indépendante de chaque facteur. Notons toutefois que les modèles ont certes l'avantage de résumer les données et de les analyser plus facilement mais ils induisent une forme d'abstraction et imposent un certain nombre de contraintes. Tout d'abord, les facteurs étudiés sont principalement ceux qui peuvent être quantifiés et facilement mesurés. Le stress, l'alimentation ou le statut socio-économique sont bien plus difficiles à mesurer que des variables physiologiques, quantitatives et

individuelles, comme l'hypertension. Dans l'enquête de Framingham, leur étude fut programmée puis abandonnée, suite aux trop grandes difficultés occasionnées. Ensuite, la forme prise par l'analyse statistique repose sur un isolement des facteurs dont on s'assure du rôle indépendant dans l'effet. C'est après seulement que leurs éventuelles interactions avec d'autres facteurs sont étudiées et prises en compte. Or cet isolement préalable conduit à négliger un certain nombre d'éléments ou de facteurs contextuels qui jouent, pour certains épidémiologistes se revendiquant d'une épidémiologie sociale ou écologique, un rôle important dans l'émergence des maladies⁶.

Ces modèles issus de la statistique mathématique furent d'abord utilisés pour *l'analyse étiologique*. Mais, rapidement, ils le furent aussi et surtout, pour la *prédiction du risque*. En particulier aux États-Unis, les cardiologues et cliniciens chercheurs, qui furent les principaux acteurs de l'épidémiologie, eurent avant tout pour objectif de se donner les moyens d'identifier le malade potentiel ou le «profil coronarien» avant que la maladie cardiovasculaire, souvent brutale, ne survienne. Les possibilités de prédiction du risque d'un individu qu'offre l'identification de plusieurs facteurs de risque pour un même individu furent attractives. Leur multiplication permet d'affiner et préciser l'identification du malade potentiel. Que les facteurs de risque soient de nature causale ou non, ils permettent de prédire le risque et ainsi de repérer le malade qui s'ignore avant que la maladie ne survienne et qu'il soit trop tard pour agir. En effet, la valeur prédictive d'un facteur n'est pas toujours identique à son importance étiologique. Les équations prédictives qui furent constituées pour calculer le risque cardiovasculaire intègrent des facteurs de risque dont le rôle n'est pas nécessairement étiologique. Par exemple, bien que le poids soit un facteur étiologique important, il se révèle peu utile pour l'équation de prédiction. Ceci conduit à souligner la différence entre *prédiction* et *analyse étiologique* et un certain primat de la visée de prédiction du risque dans l'émergence et la diffusion de la notion de facteur de risque.

Ainsi, l'étude de la genèse historique et épistémologique de la modélisation épidémiologique des facteurs de risque révèle qu'elle est fortement teintée de pragmatisme. On peut la caractériser par deux principaux traits aussi retenus par Perreti-Watel pour définir ce qu'il appelle

6. Cf. par exemple, DIEZ-ROUX (1998).

le «paradigme épidémiologique» (PERRETI-WATEL 2004)⁷: d'une part, un primat de la visée de prédiction sur l'analyse étiologique et, d'autre part, une modélisation performante de la multifactorialité. Perreti-Watel souligne que ce modèle de multifactorialité est en outre centré sur l'individu. Ceci est en effet particulièrement vrai des premiers facteurs de risque épidémiologiques identifiés pour les maladies cardiovasculaires et les cancers: ce sont principalement des attributs individuels, physiologiques ou jugés comme relevant du comportement individuel (hypertension, hypercholestérolémie, tabagisme, régime alimentaire, activité physique, etc.). Mais qu'en est-il de la causalité dans ce paradigme? Ce pragmatisme des cliniciens et des épidémiologistes est-il associé à l'abandon positiviste de la notion de cause au profit de celle de corrélation comme le préconisait notamment l'un des fondateurs de la statistique inférentielle, Karl PEARSON (1912) comme, à la même époque, le philosophe Bertrand RUSSELL (1912)?

2 ➤ L'inférence causale: le statut causal des facteurs de risque

En dépit de l'introduction de techniques de la statistique inférentielle et de son importance nouvelle dans l'analyse étiologique en épidémiologie et plus largement en médecine, il ne semble pas que le langage causal ait disparu de ces disciplines. Bien au contraire, l'identification de facteurs de risque de maladie a rapidement conduit à la question de leur statut causal. La médecine semble ne pas pouvoir s'accommoder d'une disparition de la notion de cause⁸. Cette persistance de la cause peut s'expliquer par les nécessités de l'action auxquelles la recherche étiologique est étroitement liée en médecine: à partir de quels critères convient-il de considérer qu'un facteur de risque est suffisamment important pour qu'il soit justifié de recommander et mettre en œuvre une action préventive? La question devient alors celle de l'inférence causale et de la méthodologie statistique de la preuve: est-il possible de prouver la causalité à partir d'associations statistiques? Et si ce

7. Dans cet article, Perreti-Watel dénonce entre autres l'extension de ce «paradigme épidémiologique» à l'étude des comportements et des croyances individuelles.

8. Sur la persistance de la notion de cause en médecine et de ses différentes modalités, cf. FAGOT-LARGEAULT (1993). À propos du réalisme en médecine, cf. GRENE (1976).

n'est pas le cas, quel niveau de preuve faut-il atteindre avant de pouvoir décider que la suspicion de causalité est suffisamment forte pour justifier une action préventive ?

La controverse «tabac et cancer du poumon» (1950-1964), principalement anglo-saxonne, a joué un rôle central dans l'analyse de ces questions. En effet, c'est essentiellement à l'occasion de cette controverse qu'une réflexion sur l'analyse causale en épidémiologie s'est développée et que des critères et procédures pour l'inférence ont été établis. Si dès le milieu des années 1950, plusieurs enquêtes d'épidémiologie étiologique, enquêtes cas-témoins mais aussi enquêtes de cohorte, avaient mis en évidence et établi la réalité et la solidité de la relation statistique entre la consommation individuelle de tabac et l'augmentation de risque de cancer du poumon, il n'y avait pourtant pas de consensus des chercheurs pour juger que cette relation est causale et justifie une action préventive contre le tabagisme. Les premières difficultés pour parler de facteur *causal* à propos du tabagisme reposent sur le double constat selon lequel d'une part, il y a des non fumeurs qui contractent ce cancer et des fumeurs qui ne le contractent pas (ni nécessité ni suffisance) et, d'autre part, le tabagisme est associé à de nombreuses autres pathologies que le cancer du poumon comme, par exemple, les maladies cardiovasculaires et respiratoires (absence de spécificité). Par ailleurs, il ne semblait pas y avoir de carcinogène humain en quantité suffisamment importante dans le tabac pour rendre plausible une relation causale avec le cancer du poumon. Il s'avérait donc difficile d'accepter comme causale une relation qui apparaissait ni nécessaire ni suffisante et pour laquelle on ne connaissait pas encore de mécanisme biologique bien précis. Notons que la controverse n'opposait pas les statisticiens d'un côté aux cliniciens de l'autre qui refuseraient la nouvelle méthodologie statistique de recherche étiologique. Elle divisait les statisticiens entre eux. Son enjeu était aussi la légitimation scientifique d'une nouvelle méthodologie statistique : les statisticiens qui en sont les promoteurs veulent en garantir la solidité et la protéger d'un usage médical qui pourrait manquer de rigueur (PARASCANDOLA 2004). Deux éminents statisticiens firent partie du camp des sceptiques et critiquèrent les résultats des études épidémiologiques : Ronald A. Fisher et Joseph Berkson. Pour BERKSON (1958), cette association positive observée dans diverses enquêtes pouvait être due à un biais dans la sélection des individus. Pour Fisher, un tiers facteur (un facteur génétique, en particulier) prédisposant à la fois à fumer

et à développer diverses pathologies, c'est-à-dire un biais de confusion, pouvait être en cause. Pour mettre en doute l'affirmation de causalité, Fisher s'appuyait notamment sur la trop petite différence observée dans les études entre les fumeurs qui inhalent la fumée et ceux qui n'inhalent pas. Dans les deux cas, on encourageait à chercher un mécanisme biologique sous-jacent qui expliquerait mieux car, de manière plus directe, l'association positive entre tabagisme et cancer du poumon.

La difficulté à clore la controverse conduisit les services fédéraux de santé anglais, puis américains, à demander la réalisation d'un rapport par divers experts. En 1962, le rapport anglais intitulé «Smoking and health» conclut à la causalité. Mais c'est plus particulièrement le rapport des services fédéraux américains de santé (Public Health Service) publié en 1964 qui conduisit à la normalisation du jugement de causalité et à la clôture de la controverse⁹. Ce rapport conclut aussi au statut causal de ce facteur de risque après une discussion serrée sur l'inférence causale et une section entière consacrée aux critères du jugement épidémiologique : il est affirmé qu'on ne saurait parler de «preuve» d'une relation causale mais que les données issues des études épidémiologiques couplées aux autres données réunies dans le domaine de la clinique, de la pathologie et de l'expérimentation animale fournissent une base à partir de laquelle un jugement de causalité est possible (*ibid.*, p. 20). Il est précisé que ce jugement dépasse tout énoncé de probabilité statistique. L'idée fut alors acceptée qu'il y a un lien causal et que la réduction de la consommation de tabac est bénéfique. Il s'agit donc d'un jugement et non d'une preuve et d'un jugement qui s'appuie sur un ensemble de critères d'évaluation dont aucun n'est nécessaire ou suffisant. Il est précisé que le terme de «cause» est ici utilisé dans son sens général et non pas dans son sens absolu. Cinq critères¹⁰ ont été utilisés auxquels l'épidémiologiste britannique, Bradford Hill, en ajoutera quatre (HILL 1965), livrant une liste qui devint rapidement une sorte de «boîte à outils» de l'inférence

9. Ce rapport est le résultat d'une étude réalisée par divers experts et conduite par le ministre américain de la Santé : U.S. Department of Health Education and Welfare, Surgeon General's Report (1964), *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*, Washington DC, Government Printing Office.

10. Ces cinq critères sont les suivants : la constance de l'association, sa force, sa spécificité, sa cohérence chronologique et sa plausibilité (*ibid.*, p. 20).

causale en épidémiologie: 1) la force de l'association (c'est-à-dire le fait que les enquêtes prospectives ont montré que le taux d'incidence par cancer du poumon chez les fumeurs est neuf fois plus élevé que chez les non-fumeurs); 2) la constance de cette association (le groupe d'experts auprès du ministre américain de la Santé a noté cette corrélation dans 29 études cas-témoins et 7 études prospectives de cohorte); 3) la spécificité; 4) la cohérence chronologique; 5) la relation dose-réponse ou gradient biologique; 6) la plausibilité biologique de la relation causale; 7) la cohérence avec ce que l'on sait de l'histoire naturelle de la maladie; 8) les preuves expérimentales; 9) les analogies avec d'autres pathologies. Les deux premiers critères visent à confirmer la réalité de l'association et se fondent sur des arguments statistiques. Les suivants constituent divers arguments de natures chronologique, clinique et biologique qui permettent de discuter si l'association est réellement causale. C'est donc à un faisceau d'indices hétérogènes que l'on fait appel pour décider du statut causal d'un facteur de risque. Anne FAGOT-LARGEAULT (1989, p. 373) montre que ces critères épidémiologiques de causalité sont en réalité le résultat d'une explicitation progressive de critères intuitifs qui fut encouragée par le contexte de l'élaboration et de la formalisation de méthodes statistiques d'enquête et d'analyse étiologique¹¹. Bradford Hill avait proposé ces critères dans une approche pragmatique et non dogmatique de l'inférence. Pour lui, il n'est pas nécessaire que tous soient satisfaits pour juger de la causalité d'une relation (HILL 1965, p. 299). Il insiste sur la souplesse et la nécessité de soumettre leur application à une adaptation au contexte de l'étude. Mais par ailleurs, son objectif était bien de proposer, par delà la controverse sur le tabagisme, une procédure formalisée pour l'inférence causale qui puisse valoir pour tout type de maladie. Les critères de Hill sont aujourd'hui abondamment mobilisés dans l'analyse étiologique en épidémiologie et ils sont considérés comme les critères épidémiologiques de causalité.

Parvenu à ce point, deux questions émergent au sujet de l'inférence causale et, plus généralement, du statut de la causalité en épidémiologie

11. Pour Anne Fagot-Largeault, l'explication causale en médecine relève «d'un jugement qui fait intervenir en proportion diverse selon les cas des éléments historiques (une histoire étiologique), des éléments de calcul (une inférence statistique), et des éléments décisionnels (un choix impliquant des critères de pertinence)» (2001, p. 22).

et en médecine. Une première question concerne l'aptitude de cette série de critères à valoir pour tout type de maladie. S'ils sont utiles et pertinents pour la controverse dans le contexte de laquelle ils ont été établis, valent-ils pour des pathologies comme les maladies cardiovasculaires dont les facteurs de risque ne sont pas aussi fortement associés à leur effet que le tabagisme et le cancer du poumon (le risque relatif ayant une valeur inférieure)? Les controverses sur le statut causal du cholestérol, qui n'a qu'un rôle très partiel dans les maladies cardiovasculaires, demeurent difficiles à clore. Dans ce contexte, les indices pharmacologiques et les déterminants socio-politiques mais aussi économiques et industriels interfèrent dans le jugement de causalité¹². Deuxièmement, on pourrait penser qu'en tant que «boîte à outils» de l'analyse étiologique, ces critères permettent en réalité de faire l'économie d'une réflexion sur la nature de la causalité. Pour Luc BERLIVET (1995), ils s'apparentent même à une «boîte noire» qui permettrait d'éluder la difficile question de la définition de la cause tout en répondant aux nécessités pratiques de l'inférence causale: la «déconstruction du concept en critères (relativement) faciles à appréhender offre la qualité inestimable d'être directement opératoire». Le coup de force symbolique de Hill serait en réalité d'avoir déplacé le problème «en substituant à un débat compliqué, confus, [...] une procédure formalisée de vérification un à un de neuf critères. Plus de débat philosophique, une classique opération de laboratoire s'y substitue!» (BERLIVET 1995, p. 35). Or peut-on véritablement éluder les questions relatives à la définition et à l'interprétation de la causalité sous-jacente à ces critères opérationnels? Ces questions semblent refaire surface dans les débats et avec elle, celle de l'unité, ou tout au moins, de la cohérence de la notion de causalité en médecine.

3 ⇨ L'interprétation causale

On observe en épidémiologie comme en philosophie des sciences un retour de l'analyse de la notion de cause dans ses dimensions ontologique et conceptuelle après son abandon dans la conception classique de l'empirisme logique qui réduisait la causalité à l'explication déductive-

12. Sur l'histoire de cette controverse, cf. STEINBERG (2007). Il s'agit toutefois d'une histoire qui prend partie pour la clôture de la controverse. Pour une vision philosophique et critique, cf. GREENE (2007).

nomologique¹³. Du côté de l'épidémiologie, dans un article intitulé «Causes», l'épidémiologiste américain Kenneth Rothman, auteur d'un des manuels les plus influents de l'épidémiologie moderne (ROTHMAN 1986), souligne l'importance de réduire l'écart entre les conceptions métaphysiques de la cause et les critères épidémiologiques de l'inférence (*ibid.*, p. 587). Plus récemment, Mark PARASCANDOLA & Douglas WEED (2001), philosophes et épidémiologistes au National Cancer Institute, soutiennent qu'il importe d'explicitier les critères utilisés mais aussi les définitions de la cause qu'ils présupposent, car, bien qu'implicites, ces définitions sont présentes et orientent la manière dont l'épidémiologiste analyse et modélise les phénomènes qu'il étudie.

3.1 ⇨ Le facteur de risque implique-t-il un concept probabiliste de la cause ?

Commençons par nous demander si la notion de facteur de risque renvoie à un concept particulier de causalité. Un des principaux théoriciens de l'épidémiologie moderne en France, Daniel Schwartz, lie l'usage des statistiques en médecine à une nouvelle conception de la causalité : «La révolution statistique entraîne une nouvelle conception de la cause : dans le domaine de l'aléatoire, on ne demande pas à une cause d'entraîner obligatoirement l'effet mais seulement d'en augmenter la probabilité» (SCHWARTZ 1988, p. 36). Il semblerait que l'usage de méthodologies statistiques et probabilistes conduit à l'adoption d'une conception probabiliste de la causalité en épidémiologie, conception qui constituerait une alternative à une causalité conçue en termes de nécessité et de suffisance. Des philosophes comme Hans REICHENBACH (1956), Patrick SUPPES (1970) ou Ellery EELLS (1991) ont élaboré diverses conceptions probabilistes ou statistiques de la causalité. Pour l'épidémiologiste, concevoir ainsi la

13. Le modèle de l'explication déductive-nomologique a été proposé par Carl Hempel et Paul Oppenheim dans un article de 1948 (HEMPEL & OPPENHEIM 1948). Il s'agit de concevoir l'explication scientifique comme un argument déductif concernant nécessairement au moins une loi. L'*explanandum* peut être déduit de la conjonction des deux parties de l'*explanans* : un énoncé général ou loi et des conditions initiales. Cette conception identifie l'explication scientifique à l'explication causale au sens où l'explication scientifique ainsi définie rend compte de la causalité : la cause d'un phénomène est alors l'ensemble des conditions initiales mentionnées dans une explication.

causalité implique qu'un facteur de risque est considéré comme *causal* parce qu'il augmente de façon significative la probabilité de son effet (la survenue de telle ou telle maladie). Ainsi, par exemple, un facteur qui augmente la probabilité de survenue d'une maladie cardiovasculaire chez un individu constitue une cause de maladie cardiovasculaire. Or il ne semble pas que cette conception probabiliste de la causalité soit celle adoptée par la majorité des épidémiologistes. En dépit du choix explicite de quelques uns en faveur d'un tel concept de la causalité (ELWOOD 1998, LAGIOU *et al.* 2005, PARSCANDOLA & WEED 2001), des interprétations plutôt déterministe et mécaniste semblent l'emporter. Rothman défend notamment une interprétation déterministe, ou plus précisément, une conception qui repose sur les notions de nécessité et de suffisance. Le concept, ou plutôt, le modèle défendu dans l'article précédemment évoqué (1976) est une adaptation du modèle INUS (*Insufficient but Non-redundant part of an Unnecessary but Sufficient condition*) qu'avait proposé le philosophe John MACKIE (1965) : le modèle de la « cause composée-suffisante » qui, tout en tenant compte de la multifactorialité, réintroduit les notions de nécessité et de suffisance au cœur de l'analyse causale en épidémiologie. Une « cause composée-suffisante » est en effet constituée d'un ensemble de composants dont aucun n'est par lui-même suffisant pour la survenue de la maladie mais quand l'ensemble des composants est présent, on a alors une cause suffisante. Pour Rothman, la conception probabiliste de la cause serait utile – particulièrement dans le cadre de la santé publique et quand il y a urgence à trancher avec des connaissances limitées – mais toutefois imprécise : elle ne nous dit rien de certain pour l'individu. Du côté des critiques philosophiques du concept probabiliste, Wesley SALMON (1984) soutient qu'un tel concept n'est pas suffisant pour établir la causalité ; un processus continu, conçu comme une propagation physique, doit être mis en évidence. C'est ce processus qui est déterminant pour distinguer parmi les associations statistiques celles qui sont causales.

3.2 ⇨ Les notions de cause sous-jacentes aux critères épidémiologiques de causalité : pluralisme, réductibilité ou complémentarité ?

Si l'on interprète la causalité sous-jacente aux différents critères épidémiologiques, on s'aperçoit d'une tension entre considérations de type mécaniste d'un côté et statistique ou probabiliste de l'autre, les notions

de «mécaniste» et de «probabiliste» étant ainsi entendues : «mécaniste» – sans précision sur ce qui le constitue (processus, chaînes probabilistes, dépendances contrefactuelles, etc.) – renvoyant à ce qui explique la survenue de l'effet et, «probabiliste», à ce qui introduit une différence dans l'effet (RUSSO & WILLIAMSON 2007)¹⁴. Les philosophes Federica Russo et Jon Williamson font l'analyse de ces critères épidémiologiques de causalité et montrent qu'en gros, les critères 4, 6, 7, 8, 9¹⁵ de la liste de Hill précédemment évoquée impliquent des considérations mécanistes tandis que les critères 1, 2, 5 et 8¹⁶ impliquent des considérations probabilistes. Or peut-on se satisfaire de ce dualisme ? La présence de ces deux grands type de considérations mécaniste et probabiliste pourrait être le signe d'un pluralisme causal irréductible en médecine. Mais ne s'agit-il pas plus simplement d'une suprématie déguisée des considérations mécanistes, les considérations probabilistes ne jouant en réalité qu'un rôle de repérage des hypothèses étiologiques pertinentes ? Dans le contexte de la controverse «tabac-cancer du poumon», aux deux premiers critères de type probabiliste (force de la relation et sa constance dans diverses études) que permettaient de remplir les résultats des enquêtes étiologiques dès le milieu des années 1950, il aurait fallu, pour emporter la conviction, ajouter des considérations de nature mécaniste. Les études épidémiologiques ne joueraient en réalité qu'un rôle de repérage des hypothèses étiologiques les plus pertinentes, la preuve de causalité revenant à l'analyse biologique et physiopathologique qui met en évidence des mécanismes.

Mais, si on la considère attentivement, l'histoire des controverses sur l'interprétation causale des facteurs de risque révèle plutôt la pertinence de la thèse d'une irréductible complémentarité entre considérations mécanistes et probabilistes dans le domaine médical. Pour ce qui concerne la controverse «tabac et cancer du poumon», quand les rapports concluent en 1962 puis en 1964 à la causalité, c'est surtout à

14. Il ne s'agit donc pas de la distinction entre «déterministe» et «aléatoire» mais bien de deux interprétations et définitions différentes de la notion de cause.

15. 4) cohérence chronologique, 6) plausibilité théorique, 7) cohérence, 8) les preuves expérimentales et 9) les analogies avec d'autres connaissances.

16. 1) la force de la relation, 2) sa consistance, 5) la relation dose-réponse, 8) les preuves expérimentales.

partir de critères de type statistique et probabiliste. Le jugement est certes étayé par une considération mécaniste: un mécanisme plausible de la carcinogenèse a été mis évidence. Des recherches biologiques se poursuivront ensuite pour trouver la cause première et directe, un agent qui serait certain et suffisant. La découverte en 1996 du gène P53 donne un mécanisme plus précis encore: les séquences du gène P53 qui sont mutées dans certains cas de cancer du poumon sont les mêmes que celles qui tendent à être liées à la molécule carcinogène du tabac appelée BPDE. Mais si cette découverte renforce l'inférence causale déjà acceptée, il demeure nécessaire de disposer d'informations sur l'association statistique, sa réalité et sa force, pour s'assurer de la réalité du lien. L'histoire de l'hypothèse «cholestérol» dans le risque cardiovasculaire montre un va-et-vient similaire entre considérations probabilistes et mécanistes dans le jugement de causalité. Si dès 1913, grâce à des modèles animaux, on dispose de mécanismes plausibles du rôle du dépôt de cholestérol sur la paroi artérielle dans le développement de l'athérosclérose avec les travaux de recherche expérimentale de pathologistes russes, Ludwig Aschoff (1866-1942) et Nicolai Anitschkov (1885-1964)¹⁷, c'est surtout la force et la constance des associations statistiques entre hypercholestérolémie et risque cardiovasculaire mises en évidence dans les enquêtes épidémiologiques qui contribuèrent à nourrir la conviction d'un statut causal de ce facteur de risque. En 1994, la découverte des statines¹⁸ permit de mieux comprendre les mécanismes biologiques en jeu dans cette pathologie. Mais les essais d'intervention ou essais cliniques démontrant l'efficacité des statines en termes de rapport risque et bénéfice furent tout aussi importants (STEINBERG 2007). Ainsi, les résultats statistiques obtenus dans le cadre d'enquêtes épidémiologiques d'observation sont certes insuffisants, mais la connaissance d'un mécanisme biologique l'est tout autant. Plusieurs mécanismes peuvent en effet relier un facteur à son effet présumé. Quand un mécanisme est

17. Anitschkov avait soumis des lapins à un régime alimentaire riche en œufs et la paroi vasculaire ou *intima* de ces lapins s'étaient couverte de plaques athérosclérotiques graisseuses, alors considérées comme du cholestérol (ANITSCHKOV & CHALATOW 1913).

18. Les statines forment une classe de médicaments qui, en agissant sur une enzyme de la voie du métabolisme du cholestérol, permettent de baisser le taux de LDL-cholestérol (le «mauvais cholestérol») dans le sang.

connu, l'inférence est encouragée mais il demeure nécessaire d'étudier l'association statistique et de contrôler la présence d'autres causes ou biais éventuels pour décider de la causalité car d'autres mécanismes pourraient intervenir dans la même relation entre la cause et l'effet étudiés et inversement, le même mécanisme pourrait bien conduire au même effet mais à partir d'une cause différente (THAGARD 1998). Les considérations statistiques déterminent l'étiologie, *le chemin causal*, par l'identification de la relation de cause à effet. Les considérations de type mécaniste identifient *la pathogenèse ou le processus intermédiaire* qui explique la relation de la cause à l'effet (FAGOT-LARGEAULT 1992).

La forte complémentarité de ces deux aspects de la causalité dans les sciences biomédicales nous conduit à considérer la pertinence d'une recherche de leur articulation étroite dans une conception unifiée de la cause. Plusieurs propositions ont été faites dans ce sens par des philosophes des sciences. Nous évoquons très rapidement celle de Paul THAGARD (1999) puis celle de Federica RUSSO & Jon WILLIAMSON (2007). Pour Thagard, l'explication en médecine n'est ni déductive, ni statistique, ni singulière : elle est l'«instanciation d'un réseau causal» qui caractérise plusieurs grands types de maladie (cancer, ulcère, maladies infectieuses, maladies de carence, maladies auto-immunes). L'unité est à la fois explicative et conceptuelle : elle concerne chaque grand type de maladie (THAGARD 1999, p. 20-36). Les nœuds de ce réseau sont connectés par des relations causales inférées sur la base de multiples considérations : associations, autres causes possibles, mécanismes (THAGARD 1999, p. 113-117). Par ailleurs, la constitution de ces schémas repose sur la «cohérence explicative», notion qui, bien que non analysée, était déjà présente chez certains épidémiologistes (SUSSEY 1973, ELWOOD 1988)¹⁹. De leurs côtés, RUSSO & WILLIAMSON (2007) défendent une conception épistémique à double face de la causalité. Leur objectif est de rendre compte de l'homogénéité du langage causal dans les sciences de la santé. Pour eux, l'intrication entre considérations mécanistes et probabilistes y est suffisamment forte pour justifier une recherche d'unité. Ils distinguent la *relation causale* elle-même des *genres ou types de preuve* (mécaniste et probabiliste) utilisés pour l'inférence et qui sont bien, en

19. Thagard définit cette cohérence comme une contrainte positive entre les hypothèses, telle que si l'une est acceptée, l'autre aussi et inversement (THAGARD 1999, p. 64-69).

effet, d'une dualité irréductible dans les sciences de la santé. Il peut bien y avoir deux genres de preuve pour l'assertion causale et pourtant une unique relation de cause. Pour eux, la relation de cause n'est pas une réalité ontologique, elle est épistémique : elle relève d'une croyance rationnelle. Ainsi, la dualité des types de preuves (mécanistes et probabilistes) est maintenue au sein d'une conception épistémique unifiée de la causalité. Dans cette configuration, le lien est étroit entre l'analyse conceptuelle et ontologique de la cause et l'analyse épistémologique. La pertinence et la validité de cette démarche restent à examiner. Ce domaine de recherche est en plein essor et place la médecine, et en particulier l'épidémiologie, au cœur d'une réflexion fondamentale pour la philosophie des sciences.

* * * * *

Pour conclure, notre analyse historique et épistémologique du rapport de la notion de facteur de risque avec celle de cause nous a conduit à mettre en évidence le rôle important de l'épidémiologie dans le développement de procédures pour l'inférence causale mais aussi dans le renouveau de la réflexion sur le concept de cause. Aux frontières entre sciences sociales et sciences biologiques, l'épidémiologie et ses méthodes de recherche sur la causalité sont l'occasion d'une réflexion abondante et féconde sur ce sujet si central pour la philosophie des sciences. Une convergence prometteuse se fait jour entre la réflexion des épidémiologistes et des statisticiens et celle des philosophes pour renouveler et approfondir l'analyse de la causalité.

☉Références bibliographiques☾

A

ANITSCHKOV Nicolai & CHALATOW Sergejewitsch S. (1913), "Über experimentelle Cholesterinsteatose und ihre Bedeutung für die Entstehung einiger pathologischer Prozesse", *Zentralblatt für allgemeine Pathologie und pathologische Anatomie*, 24: 1-9.

ARONOWITZ Robert (1999), *Les Maladies ont-elles un sens ?* [1998], Synthélabo.

B

BERKSON Joseph (1958), "Smoking and lung cancer: Some recent observations on two reports", *Journal of the American Statistical Association*, 53: 28-38.

BERLIVET Luc (1995), *Controverse en épidémiologie. Production et circulation de statistiques médicales*, rapport pour la Mire, Rennes, CNRS.

C

CORNFIELD Jerome (1951), "A method of investigating comparative rates from clinical data. Application to cancer of the lung, breast, cer-vix", *Journal of the National Cancer Institute*, 11: 1269-1275.

CORNFIELD Jerome (1962), "Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis", *Federation Proceedings*, 21: 58-61.

D

DIEZ-ROUX Ana (1998), "Bringing the context back into epidemiology", *American Journal of Public Health*, 88: 216-222.

DOLL Richard & PETO Richard (1976), "Mortality in relation to smoking: 20 years' observations on male British doctors", *British Medical Journal*: 1525-1536.

E

ELLS Ellery (1991), *Probabilistic Causality*, Cambridge University Press.

ELWOOD Mark J. (1988), *Causal Relationships in Medicine*, Oxford Medical Publication.

EVANS Alfred S. (1993), *Causation and disease: A chronological journey*, Plenum.

F

FAGOT-LARGEAULT Anne (1989), *Les Causes de la mort, histoire naturelle des facteurs de risque*, Vrin.

FAGOT-LARGEAULT Anne (1992), «Quelques implications de la recherche étiologique», *Sciences sociales et santé*, 10: 33-45.

FAGOT-LARGEAULT Anne (1993), "On medicine's scientificity – did medicine's accession to scientific 'positivity' in the course of nineteenth century require giving up causal (etiological) explanation?" in C. Delkeskamp-Hayes & M.A. Gardell Cutter (eds.), *Science, Technology and the Art of Medicine*, Kluwer Academic Publishers: 105-126.

FAGOT-LARGEAULT Anne (2001), «Leçon inaugurale du 1^{er} mars 2001 au Collège de France», chaire de philosophie des sciences biologiques et médicales, Collège de France.

FISHER Ronald A. (1935), *The Design of Experiments*, Oliver and Boyd.

FISHER Ronald A. (1959), *Smoking: The Cancer Controversy*, Oliver and Boyd.

G

GIROUX Élodie (2006), *Épidémiologie des facteurs de risque: genèse d'une nouvelle approche de la maladie*, thèse de doctorat en philosophie de la médecine, Paris, université de Paris 1 Panthéon-Sorbonne.

GIROUX Élodie (2008), «Enquête de cohorte et analyse multivariée: une analyse épistémologique et historique du rôle fondateur de l'étude de Framingham», *Revue d'épidémiologie et de santé publique*, 56: 177-188.

GREENE Jeremy A. (2007), *Prescribing by Numbers: Drugs And the Definition of Disease*, The Johns Hopkins University Press.

GRENE Marjorie (1976), "Philosophy of Medicine: Prolegomena to a Philosophy of Science", *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*: 77-93.

H

HEMPEL Carl G. & OPPENHEIM Paul (1948), "Studies in the Logic of Explanation", in C.G. Hempel, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, Free Press, 1965.

HILL Bradford A. (1965), "Environment and disease: association or causation?" *Proceedings of the Royal Society of Medicine* 58: 295-300.

K

KANNEL William B., DAWBER Thomas, KAGAN A., REVOTSKIE N. & STOKES J.I. (1961), "Factors of risk in the development of coronary heart disease, six-year follow-up experience, the Framingham study", *Annals of Internal Medicine*, 55: 33-48.

L

LAGIOU P., ADAM H.O. & TRICHOPOULOS D. (2005), "Causality, in cancer epidemiology", *European Journal of Epidemiology*, 20: 565-574.

M

MACKIE John L. (1965), "Causes and conditions", *American Philosophical Quarterly*, 2: 245-264.

MACMAHON Brian, PUGH Thomas F. & IPSEN Johannes (1960), *Epidemiologic Methods*, J. & A. Churchill.

O

OPPENHEIMER Gerald M. (2006), "Profiling risk: the emergence of coronary heart disease epidemiology in the United States (1947-70)", *International Journal of Epidemiology*, 35, 3: 720-730.

P

PARASCANDOLA Mark (2004), "Skepticism, statistical methods, and the cigarette: a historical analysis of a methodological debate", *Perspectives in Biology and Medicine*, 47: 246-261.

PARASCANDOLA Mark & WEED Douglas (2001), "Causation in epidemiology", *Perspectives in Biology and Medicine*, 55: 905-912.

PEARSON Karl (1912), *La Grammaire de la science* [1892], Alcan.

PERRETI-WATEL Patrick (2004), «Du recours au paradigme épidémiologique pour l'étude des conduites à risque», *Revue française de sociologie* : 103-132.

R

REICHENBACH Hans (1956), *The Direction of Time*, University of California Press.

ROTHMAN Kenneth J. (1976), "Causes", *American Journal of Epidemiology*, 104: 587-592.

ROTHMAN Kenneth J. (1986), *Modern Epidemiology*, Little, Brown and Company.

RUSSELL Bertrand (1912), "On the notion of cause", *Proceedings of the Aristotelian Society*, 13: 1-26. Trad. fr., «Sur la notion de cause», *Philosophie*, n° 89, printemps 2006: 3-20.

RUSSO Federica & WILLIAMSON Jon (2007), "Interpreting causality in the health sciences", *International Studies in the Philosophy of Science*, 21: 157-170.

S

SALMON Wesley C. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton University Press.

SCHWARTZ Daniel (1988), «L'irrésolu», in J. Lellouch (dir.), *Présent et futur de l'épidémiologie*, Inserm.

SKOLBEKKEN John-Arne (1995), "The risk epidemic in medical journals", *Social Sciences and Medicine*, 40, 3: 291-305.

STEINBERG Daniel (2007), *The Cholesterol Wars: The Skeptics Versus the Preponderance of Evidence*, Academic Press.

SUPPES Patrick (1970), *A Probabilistic Theory of Causality*, North-Holland Publishing Co.

SUSSER Mervin (1973), *Causal Thinking in the Health Sciences*, Oxford University Press.

T

THAGARD Paul (1998), "Explaining Disease: Correlations, Causes, and Mechanisms", *Minds and Machines*, 8: 61-78.

THAGARD Paul (1999), *How Scientists Explain Disease*, Princeton University Press.

☞ Notice biographique ☞

Élodie Giroux est maître de conférences en philosophie des sciences au département de philosophie de l'Université Jean-Moulin Lyon 3 et membre de l'Institut de recherches philosophiques de Lyon. Ses travaux en philosophie de la médecine portent sur l'histoire et l'épistémologie de l'épidémiologie des facteurs de risque, la définition de la santé et de la maladie et l'analyse causale en épidémiologie et en médecine. Dernières publications : « Enquête de cohorte et analyse multivariée : une analyse épistémologique et historique du rôle fondateur de l'étude de Framingham », *Revue d'épidémiologie et de santé publique*, 2008, 56, 3, 177-188 ; « Définir objectivement la santé : une évaluation du concept bio-statistique de Boorse à partir de l'épidémiologie moderne », *Revue philosophique*, 2009, 134, 1, 35-58 ; « Après Canguilhem : définir la santé et la maladie », *Philosophies*, PUF, 2010 ; « Philosophie de la médecine », in A. Barberousse, M. Cozick & D. Bonnay, *Précis de philosophie des sciences*, Vuibert, à paraître en 2010. <elodie.giroux@univ-lyon3.fr>.

Avertissement. Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur et des auteurs. Les textes et illustrations figurant sur ce site peuvent être consultés et reproduits sur un support papier ou numérique sous réserve qu'ils soient strictement réservés à un usage personnel, scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document selon ce modèle :

Élodie Giroux, « Facteurs de risque et causalité en épidémiologie », in Gérard Lambert et Marc Silberstein (dir.), *Matière première. Revue d'épistémologie [en ligne]. Nouvelle série, N° 1/2010 : Épistémologie de la médecine et de la santé*, Éditions Matériologiques. Mis en ligne le 7 octobre 2010. URL : www.materiologiques.com

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.





Steeves Demazeux Le concept de fonction dans le discours psychiatrique contemporain

Au milieu des années 1970, on observe parallèlement, dans deux champs disciplinaires en train de se constituer, la philosophie de la biologie d'une part et la philosophie de la psychiatrie d'autre part, une passion soudaine pour le vieux concept de fonction. Du côté de la biologie, la dimension téléologique du concept a toujours posé un problème particulier au philosophe. Dans un article qui va nourrir le débat pour les décennies à suivre, Larry Wright, en 1973, tente de résoudre la difficulté en proposant une conception étiologique de la fonction qui ramène les caractéristiques téléologiques apparentes de nos attributions fonctionnelles à une forme d'explication causale ordinaire (WRIGHT 1973). Robert Cummins, en 1975, lui oppose une conception de la fonction qu'on peut qualifier de « systémique », qui se veut plus générale mais aussi plus fidèle à l'usage traditionnel des biologistes et des médecins. Selon lui, une fonction désigne la contribution causale d'un élément à l'intérieur d'un système (CUMMINS 1975). Du côté de la psychiatrie, le problème touche d'abord à la notion de « maladie mentale ». Depuis les années 1960, nombreux sont les philosophes, les sociologues, les médecins eux-mêmes à dénoncer l'usage abusif qui est fait des étiquettes psychiatriques pour corriger toutes sortes de comportements déviants. La psychiatrie en tant que discipline médicale est en crise, et l'une des questions les plus brûlantes est de savoir si ses bases conceptuelles sont légitimes, à commencer par le concept général de « trouble » ou de « maladie » mentale autour duquel son discours s'organise. L'idée

que le cœur du problème résiderait dans le concept de dysfonctionnement – et corrélativement dans celui de fonction – est ancienne, mais elle peine à être clairement défendue. Christopher Boorse, en 1975, à la croisée des réflexions biologique et médicale, propose une analyse fine du concept de maladie mentale (*mental disease*) qu'il fait reposer sur celui de (dys)fonctionnement mental. Jerome Wakefield, plus tard, en 1992, lui emboîte le pas avec une analyse qui s'appuie plus explicitement sur une conception évolutionniste des fonctions mentales.

Nous nous proposons de revenir sur les approches fonctionnelles respectives de ces deux auteurs afin d'en formuler une synthèse critique. Mais au rebours de la plupart des commentaires ou des critiques qui se sont placés du point de vue de la biologie et de la philosophie de l'esprit, ou alors d'un point de vue médical très large¹, nous chercherons à examiner leur contribution à la lumière particulière des solutions qu'ils proposent pour la psychiatrie. De fait, l'un comme l'autre, Boorse et Wakefield ont clairement affiché leur prétention d'apporter à la psychiatrie une analyse susceptible de l'aider à sortir de ses ambiguïtés conceptuelles et de son marasme épistémologique. Mais cette prétention des deux auteurs n'est pas elle-même sans ambiguïté : les solutions respectives qu'ils proposent hésitent entre deux points de vue contradictoires. Celui d'une dénonciation de ce que le discours psychiatrique est au regard de ce qu'il devrait être. Celui à l'opposé d'une légitimation de ce qu'il essaie d'être au regard de ce qu'il ne peut qu'être dans l'état actuel des connaissances. Nous nous proposons de montrer que cette hésitation, qui tient selon nous au rôle trop indéterminé que les deux auteurs font jouer, chacun à sa manière, au concept de fonction biologique dans leur définition du trouble mental, compromet l'intérêt pratique que l'on peut tirer des analyses de Wakefield et de Boorse.

1. Nous renvoyons particulièrement aux thèses de Karen Neander, *Abnormal psychobiology* (PhD Thesis, La Trobe University, Australie, 1983) et de Marie-Claude Lorne, *Explications fonctionnelles et normativité : analyse de la théorie du rôle causal et des théories étiologiques de la fonction* (thèse de philosophie, École des hautes études sciences sociales, Paris, 2004), ainsi qu'à l'ouvrage de Mahesh Ananth, *In Defense of an Evolutionary Concept of Health: Nature, Norms, and Human Biology* (Ashgate Studies in Applied Ethics, 2008).

1 ↻ Le programme d'une analyse conceptuelle du concept de maladie

1.1 ↻ Le contexte polémique

Les contributions de Boorse et de Wakefield à la philosophie de la médecine ont assurément une portée générale. Leur prétention respective n'est autre que de résoudre le vieux problème de la démarcation entre le normal et le pathologique pour l'ensemble des phénomènes médicaux. Pourtant, le problème initial auquel ils se confrontent est précis et spécifique : il s'agit de répondre aux critiques antipsychiatriques qui considèrent, pour le dire brièvement, que la médecine mentale est une médecine usurpée. Boorse et Wakefield, dans leurs travaux respectifs, entendent montrer que la psychiatrie n'est pas une médecine usurpée, ou du moins – et c'est toute l'ambiguïté qu'on veut soulever – qu'elle n'est pas nécessairement amenée à l'être. Leur ambition première, c'est de fournir une définition objective (partiellement objective chez Wakefield) du trouble mental. En un sens, si leur travail a une portée générale, c'est d'abord parce que la validité générale de la définition qu'ils proposent du phénomène pathologique sera considérée comme la meilleure garantie qu'on puisse apporter de sa validité locale à l'échelle de la psychiatrie. Autrement dit, les deux auteurs s'accordent dès le départ pour considérer que la « maladie mentale » (*mental disease*) doit être définie comme une espèce particulière à l'intérieur du genre « maladie » (*disease*) commun à l'ensemble du discours médical.

Dans ce qui suit, nous ne nous intéresserons qu'à l'intention précise et première qu'ont eue ces deux auteurs de répondre aux critiques antipsychiatriques. Si Wakefield a toujours explicitement inscrit son travail dans le cadre d'une réflexion sur la psychiatrie, on a pu reprocher à Boorse, occupé à répondre aux objections générales qui lui ont été adressées, de négliger finalement le problème précis dont il était parti². Pourtant, l'auteur lui-même, dans un texte de synthèse de 1997, regrette le peu d'attention portée par les critiques de sa théorie au « but

2. Un tel reproche a été soulevé par William Fulford dans une belle lettre ouverte à Christopher Boorse. En substance, Fulford regrette que Boorse se soit trompé de perspective, dans l'idée que ce n'est pas par une meilleure compréhension de la maladie en générale qu'on pourra mieux comprendre la spécificité de la maladie psychiatrique ; mais que c'est au contraire par une meilleure compréhension de la spécificité de la maladie psychiatrique

principal» de celle-ci, qui était bien de proposer une analyse renouvelée de la santé mentale (BOORSE 1997). Un rappel du contexte polémique qui a servi de ferment à la réflexion des deux auteurs s'impose donc. Le premier article de Boorse sur la question (BOORSE 1975), le plus cité, date de 1975 et s'inscrit explicitement dans le débat antipsychiatrique de l'époque. Thomas Szasz, le fameux et très controversé psychiatre d'origine hongroise appelé à devenir le «pape» de l'antipsychiatrie américaine, a décoché ses premières flèches contre l'institution psychiatrique. Sa démonstration, suivant laquelle les maladies mentales ne sont rien d'autre que des «mythes» (cf. SZASZ 1960), a fait mouche: nonobstant le caractère pamphlétaire et souvent péremptoire de ses textes, Szasz pointe de réelles difficultés épistémologiques. Boorse, dans son article, le cite dès l'introduction, et s'accorde avec lui pour reconnaître que la légitimité de l'extension du vocabulaire médical à la sphère des états mentaux est à tout le moins *une question ouverte* (BOORSE 1975, p. 50). Il reconnaît d'ailleurs le «tournant psychiatrique» comme un fait caractéristique du *xx*^e siècle, dans le sens d'une tentation grandissante de régler des problèmes sociaux et moraux en s'appuyant sur le paradigme de la médecine. Mais l'enjeu du texte de Boorse de 1975, ainsi que de ceux qui suivront en 1976 et 1977, est d'apporter une réponse positive à cette question ouverte: assurément, le projet d'une médecine proprement «mentale» fait sens, si tant est seulement qu'on soit en mesure de définir un ensemble de fonctions mentales typiques de l'espèce humaine.

Le travail de Wakefield est plus tardif. Ses premiers articles consacrés à la définition du trouble mental datent du début des années 1990, et ils s'inscrivent dans un contexte polémique certes apaisé, mais loin d'être pacifié. Reprochant à la définition qui surplombe la troisième version du *Manuel diagnostique et statistique des troubles mentaux* (le DSM-III) d'être ambiguë et trop laxiste, Wakefield propose une «analyse du dysfonctionnement préjudiciable» (*Harmful Dysfunction Analysis*, HDA), qui cherche à cerner les deux pôles fondamentaux devant selon lui ordonner toute caractérisation d'un trouble mental: un pôle factuel, qui repose sur la reconnaissance spécifique d'un certain dysfonctionnement mental, et un pôle appréciatif ou normatif, qui exige de ce dys-

qu'on sera amenés à réformer notre conception de la maladie en général. Cf. FULFORD (2001, p. 80-85).

fonctionnement qu'il soit perçu comme «préjudiciable» (*harmful*) au regard des normes sociales. Ces deux éléments – le dysfonctionnement mental d'un côté (*dysfunction*), le caractère préjudiciable (*harmful*) de ce dysfonctionnement de l'autre côté – sont suivant la HDA les deux conditions nécessaires et suffisantes pour que l'on puisse légitimement caractériser un «trouble mental». Comme Boorse, il s'agit avant tout de s'opposer aux antipsychiatres qui ne voient dans l'étiquette de «trouble mental» qu'un mythe ou au mieux l'expression d'un pur jugement de valeur. Mais comme lui, il s'agit aussi de s'opposer aux comportementalistes, qui s'accommodent fort bien de la charge normative du concept, et qui n'hésitent pas à réduire le «trouble mental» à un «trouble du comportement», quitte ainsi à réduire le phénomène pathologique à un problème d'adaptation sociale. Autrement dit, chez ces deux auteurs, le concept de fonction va jouer un rôle discriminant tout à fait essentiel pour distinguer le normal du pathologique.

1.2 ↻ Deux versions modernes du principe de Broussais

Boorse et Wakefield ne sont évidemment pas les premiers à tenter de promouvoir une approche objectiviste du phénomène pathologique dans sa généralité. Canguilhem, dans *Le Normal et le pathologique*, a consacré l'expression de «principe de Broussais» pour désigner cette vieille aspiration à appréhender dans une lumière pleinement positive le phénomène pathologique, en l'identifiant dans sa réalité factuelle au défaut ou à l'excès d'un phénomène physiologique. Il emprunte l'expression à Auguste Comte, qui avait vanté, dans la 40^e leçon de son *Cours de philosophie positive*, les mérites du médecin François Broussais (1772-1838) d'avoir su le premier clairement formuler ce principe d'une continuité essentielle entre les phénomènes physiologiques et les phénomènes pathologiques : «Jamais, écrit-il, on n'a conçu d'une manière aussi directe et aussi satisfaisante la relation fondamentale entre la pathologie et la physiologie» (cité par CANGUILHEM 1998, p. 18). On connaît la critique attentive que Canguilhem fournit de ce «principe de Broussais», qui manque selon lui ce que le phénomène pathologique a pour la médecine d'originellement irréductible, dans la mesure où il néglige l'importance de la normativité inhérente au vivant. Contentons-nous de relever les deux objections dirimantes que Canguilhem adresse au principe de Broussais :

L'ambition de rendre la pathologie et par suite la thérapeutique intégralement scientifiques, en les faisant procéder simplement d'une phy-

siologie préalablement instituée, n'aurait de sens que si d'abord une définition purement objective pouvait être donnée du normal comme d'un fait, si de plus on pouvait traduire toute différence entre l'état normal et l'état pathologique dans le langage de la quantité, car la quantité seule peut rendre compte à la fois de l'homogénéité et de la variation. (CANGUILHEM 1998, p. 26.)

Ce que Canguilhem a voulu démontrer, c'est que ni Broussais ni Comte, ni Bernard plus tard ne sont parvenus à remplir ces deux exigences fortes. Boorse et Wakefield, ces deux défenseurs d'une version moderne du principe de Broussais, feraient-ils mieux? Quitte à prendre un peu de liberté avec le texte canguilhémien, reformulons et formalisons ces deux clauses du contrat objectiviste. Elles nous serviront de points d'appui tout au long de notre argumentation.

Une caractérisation objective du phénomène pathologique ne peut être obtenue que si :

- (1) une définition purement factuelle du normal est proposée ;
- (2) la différence entre le normal et le pathologique peut être traduite d'une manière quantitative.

Concernant la première clause du contrat objectiviste (1), on s'accordera sans difficulté à penser qu'il s'agit d'une clause à laquelle aucune tentative de définition objective du phénomène pathologique ne saurait échapper. L'enjeu consiste à proposer du normal une caractérisation qui ne le cantonne pas à un simple jugement de valeur. L'accent est mis sur la pertinence théorique d'une définition factuelle du phénomène normal. Or on peut déjà remarquer que Boorse et Wakefield entendent satisfaire cette exigence de la même manière. D'abord sur le plan de la méthode : nos deux auteurs s'appuient sur la méthode dite de l'*analyse conceptuelle*, procédé commun dans la philosophie anglo-saxonne, pour tenter de clarifier et de mettre en évidence la pointe factuelle du concept de maladie (*disease*) – à savoir la référence à un certain *dysfonctionnement*. Ensuite, sur le plan du résultat : nos deux auteurs insistent sur le fait que c'est grâce au concept de *fonction biologique* qu'on sera à même de parler du normal comme d'un fait. On peut décrire la normalité d'une fonction en toute objectivité, dès lors simplement que l'on sait en quoi consiste cette fonction. Dire par exemple d'un cœur qu'il fonctionne *normalement*, cela peut être appréhendé sur le plan biologique, et contrairement aux apparences, comme un jugement descriptif et factuel. Tel est l'enjeu de la première clause.

La seconde clause (2) est plus délicate à interpréter. Pour Canguilhem, la différence entre le normal et le pathologique peut être conçue soit comme une différence qualitative (une différence de nature), soit comme une différence quantitative (une différence de degré). Si, dans la pensée de Canguilhem, la polarité dynamique du vivant doit nous amener à penser cette différence comme une différence qualitative (Canguilhem défendra l'idée que la maladie pour le malade est avant tout «une autre allure de la vie» [*ibid.*, p. 51]), le seul moyen en revanche de penser cette différence comme objective serait de pouvoir la «traduire dans le langage de la quantité», «car la quantité seule peut rendre compte à la fois de l'homogénéité et de la variation» (*ibid.*, p. 26). Il s'agit là d'un argument très général qui s'intéresse aux conditions de possibilité pour penser une identité malgré la différence, et une différence à travers l'identité. Comment en effet penser la pathologie comme obéissant aux mêmes règles générales que la physiologie, tout en invoquant une différence objective entre les deux disciplines? Sur un plan plus pratique, cela revient à poser la question suivante: comment *identifier* le phénomène pathologique de manière objective parmi tous les phénomènes du vivant? Tandis que la clause (1) porte plutôt sur la question de la légitimité conceptuelle de l'approche objectiviste – sa justification en raison –, la clause (2) porte davantage sur les moyens de traduire cette légitimité dans les faits ou, pour le dire autrement, sur les conditions de son application pratique dans les sciences biologiques et médicales.

Revenons à Boorse et Wakefield, ainsi qu'à la psychiatrie. La grande pertinence des deux approches, pensons-nous, a été de proposer une solution intéressante au problème de la légitimité conceptuelle de l'approche objectiviste en psychiatrie (1). Contre Szasz, qui avait fondé son attaque sur le rôle fédérateur du concept de *lésion* en médecine somatique, la réévaluation du concept de *fonction biologique* découvre le rôle potentiellement structurant que celui-ci peut avoir dans le discours médical en général, et dans le discours psychiatrique en particulier³. Cette réévaluation permet ainsi de défendre sur de nouvelles bases la

3. Il faut bien voir en ce sens que Boorse et Wakefield, en réhabilitant à travers le concept de fonction l'autonomie du discours psychologique en médecine, sont aux antipodes de Broussais qui n'avait de cesse, dans *De l'irritation et de la folie*, de s'en prendre aux «psychologistes» – lesquels ne pouvaient être à ses yeux que des spiritualistes déguisés. Mais nous nous intéressons ici

légitimité du modèle médical en psychiatrie, et donc la légitimité de la psychiatrie comme authentique branche du savoir médical.

Néanmoins, c'est une chose de montrer que la psychiatrie *peut* être une discipline médicale légitime s'occupant de conditions qui mériteraient vraiment l'étiquette de «maladies mentales». Mais cela en est une autre pour le philosophe que d'adouber, en réponse à ceux qui le critiquent, le discours psychiatrique tel qu'il *est*, dans ses usages et dans ses pratiques. Car suffit-il qu'un discours soit cohérent et pertinent dans son projet théorique pour laver de tout soupçon sa mise en pratique? Ce serait là grande naïveté de la part d'un philosophe. L'histoire, même scientifique, est pleine de ces horreurs où l'on se prévaut d'une légitimité théorique pour justifier des pratiques distordues. À cet égard, il n'est pas vain de rappeler que les critiques antipsychiatriques, dans leur grande majorité, se sont le plus souvent attaquées au discours psychiatrique tel qu'il se présente dans la réalité plutôt que tel qu'il pourrait ou voudrait être. Szasz, avec son objection transcendantale (au sens où il s'agit d'une objection qui s'attaque aux conditions de possibilité mêmes du discours psychiatrique), est plutôt l'exception que la règle. Le plus souvent, quand ce n'est pas au pouvoir insidieux exercé par l'institution psychiatrique, c'est aux étiquettes particulières que les critiques de la psychiatrie s'en prennent, en remettant en cause leur construction théorique ou en interrogeant le pouvoir discriminant qu'on leur prête de distinguer entre des comportements normaux et des comportements pathologiques. Autrement dit, plutôt que le problème de la légitimité globale (1), c'est celui des moyens d'une démarcation spécifique qui véritablement importe (2) : quelles sont les preuves, quelles sont les raisons effectives qui permettent de distinguer tel comportement sain de tel comportement pathologique? Sur quoi la psychiatrie se fonde-t-elle pour opérer dans la pratique son découpage nosologique?

1.3 ↻ L'attitude du philosophe devant la psychiatrie de son époque

Devant ce qu'on appelle ici le problème de la démarcation spécifique du normal et du pathologique (2), il faut donc bien distinguer entre la manière effective dont s'y prend le discours psychiatrique institué, et la manière dont, peut-être, la psychiatrie devrait s'y prendre. Or on peut

au «principe de Broussais», tel qu'il a été caractérisé par Comte et approfondi par Canguilhem.

dire que Boorse et Wakefield adoptent une même attitude générale face à ce problème, bien qu'ils y apportent, comme nous le verrons, une réponse sensiblement différente. Leur attitude générale est marquée par deux traits récurrents.

Le premier trait, c'est la prudence à l'égard du discours psychiatrique institué. Boorse, à plusieurs reprises, prend soigneusement ses distances avec la psychiatrie de son époque. Il reconnaît volontiers qu'il y a un abus dans les étiquettes psychiatriques, tout au moins une certaine légèreté de la profession à qualifier de pathologiques des comportements que la société réprouve (BOORSE 1976, p. 75). Le titre de l'article de 1976, "What a theory of mental health should be", montre bien en ce sens que son ambition est davantage de réformer ou d'assainir le projet psychiatrique que de défendre une discipline instituée qui serait en souffrance. La conclusion de cet article est d'ailleurs éloquente, puisque l'auteur a l'impression que la psychiatrie de son époque se trouve à la « croisée des chemins » (*ibid.*, p. 81), et qu'elle doit faire un pas vers plus de clarté et de rigueur théoriques, ou alors renoncer avec franchise à toute référence au normal et au pathologique. Mais, Boorse tempère-t-il aussitôt, cela ne signifie pas qu'il faille être radicalement défiant à l'encontre du savoir psychiatrique en l'état : malgré les disputes et la confusion, l'auteur imagine mal que les cliniciens en général soient – pour emprunter une expression familière qui traduit littéralement son idée – « complètement à côté de la plaque » (*too wide of the mark*) (*ibid.*). Wakefield, dans son travail, partagera la même attitude qu'on peut qualifier avantageusement de *critique constructive* à l'égard de la psychiatrie instituée. Avec pour sa part assurément plus d'enthousiasme et d'engagement dans les débats particuliers. Wakefield est un auteur prolifique, qui croit dans le bien-fondé de la plupart des étiquettes diagnostiques, mais qui entend rester très sensible aux pouvoirs normatifs dont elles sont souvent l'objet et l'enjeu⁴. Son analyse a porté sur de nombreuses questions plus ou moins litigieuses, des troubles de la personnalité au diagnostic de la dépression, des troubles anxieux aux troubles sexuels. S'il fallait résumer sa position, on pourrait dire qu'elle consiste dans une défense

4. Wakefield fut, soit dit en passant, un étudiant de Foucault à Berkeley. Même s'il a rapidement pris ses distances avec les positions antifinaturalistes du philosophe français, celui-ci demeure une référence importante dans son travail.

appuyée du projet du DSM-III (et de ses versions ultérieures) mâtinée d'une critique souvent incisive portant sur les différents aspects de sa réalisation pratique.

Il est un deuxième trait commun, concernant la psychiatrie instituée, qui mérite d'être indiqué au croisement des deux argumentations. C'est celui qui consiste à souligner le fait que la psychiatrie scientifique est encore en enfance. À l'instar de Claude Bernard en son temps qui regrettait que la médecine somatique soit encore « une science dans l'enfance », contrairement à ces « sciences constituées » que sont par exemple la physique ou la chimie, mais qui ne voyait pas en cela une raison d'être attentiste, considérant au contraire que l'« on gagne souvent à expérimenter » (BERNARD 1984 [1965], p. 50), Boorse et Wakefield font valoir que la psychiatrie de notre époque doit proposer des solutions avec les moyens dont elle dispose, en attendant de pouvoir s'appuyer positivement sur la « physiologie des fonctions mentales » dont elle a besoin. Or chez les deux auteurs, l'approche objectiviste qui est préconisée, au creux même de cette situation épistémologiquement insatisfaisante, est présentée comme fructueuse et opérante. Boorse vante la « richesse conceptuelle » et la « flexibilité pratique » (BOORSE 1997, p. 101) de sa théorie, considérant que « dès lors qu'un dysfonctionnement biologique est requis pour qu'il y ait trouble [pathologique], pratiquement tous les avantages de la BST [*Bio-Statistical Theory*]⁵ concourent pour clarifier les controverses professionnelles et sociales et pour prévenir l'abus politique du vocabulaire médical » (*ibid.*, p. 100). Quant à Wakefield, il a constamment mis en avant la « puissance explicative considérable de la HDA pour comprendre la distinction entre un trouble mental et tout autre état mental problématique » (WAKEFIELD 2007, p. 149), mettant à plusieurs reprises cette HDA en pratique pour tester la validité des critères diagnostiques de certains troubles mentaux.

Entre prudence critique et marques de confiance adressées au discours psychiatrique de leur époque, Boorse et Wakefield se sont faits les avocats d'une médecine scientifique dont il y a peu de raisons de douter qu'elle soit au service des patients plutôt qu'au service du pouvoir insidieux et normalisateur de l'institution psychiatrique. Leur conviction commune, c'est simplement que la science et la clarté conceptuelle peuvent déminer un certain nombre de questions litigieuses en psy-

5. C'est l'étiquette que Boorse donnera à sa théorie. Cf. *infra*.

chirurgie. On serait convaincu à moins. Seulement voilà, sont-elles en mesure de le faire ?

2 ↻ L'approche populationnelle de Boorse

2.1 ↻ La psychiatrie, une entreprise théorique légitime

Boorse peut être perçu comme celui qui répond le plus diligemment au défi fixé par Canguilhem : il entend (1) proposer une définition purement factuelle du normal, qui s'appuie sur une conception non normative (au sens de non polluée par des jugements de valeur) de la fonction en biologie ; il prétend pouvoir (2) traduire la différence entre le normal et le pathologique d'une manière quantitative, au moyen d'une approche qu'il qualifiera de « théorie bio-statistique » (BST).

Le premier point (1) est celui qui a le plus retenu l'attention des philosophes : peut-on définir objectivement le phénomène pathologique sur le constat d'un dysfonctionnement biologique ? La solution proposée par Boorse en 1975 est simple et élégante, mais elle soulève de nombreuses difficultés d'ordre général. Dans un article de synthèse écrit en 1997, "A Rebuttal on Health", l'auteur a répondu avec un soin attentif aux nombreuses objections techniques, médicales et biologiques qui lui ont été adressées en l'espace de vingt ans. On s'étonne néanmoins de constater que dans cet article qui s'étend sur plus de cent pages, les références spécifiques au problème psychiatrique, si on les regroupait pour les mettre bout à bout, n'excéderaient pas dix pages⁶. Le problème psychiatrique, qui était l'un des enjeux majeurs de ses articles médicaux des années 1970, est devenu complètement secondaire dans le texte de 1997⁷. La raison d'un tel déplacement est évidente : pour défendre la cohérence générale de son projet théorique, Boorse a naturellement été conduit à privilégier de plus en plus les exemples percutants et intuitifs tirés de la médecine somatique au détriment de ceux, souvent équivoques et polémiques, qu'il aurait pu tirer de la psychiatrie. Il n'empêche

6. Dont une grande partie consacrée à une discussion avec Engelhardt sur la masturbation au XIX^e siècle.

7. On peut néanmoins deviner qu'il demeure, en 1997, l'horizon lointain de toutes les discussions. Il suffit pour s'en convaincre de remarquer que la plupart des auteurs discutés sont directement impliqués dans le champ de la psychiatrie. C'est le cas par exemple de Fulford, de Reznek, de Kendell, de Klein, de Scadding, ou de Wakefield.

qu'au vu de l'importance donnée à des objections de plus en plus fines, voire tatillonnes (une seule cellule qui dysfonctionne, cela suffit-il à faire une maladie? Que penser de ces dysfonctionnements qui protègent de certaines maladies, comme l'hémoglobine HbS pour le paludisme? etc.), on ne peut que déplorer le fait que la priorité désormais ne soit plus d'offrir à la psychiatrie les moyens théoriques de répondre aux objections qui lui sont faites, mais simplement de défendre la cohérence d'une théorie générale pour elle-même. Comme cela arrive souvent, la discussion philosophique, toute occupée à disputer scolastiquement des difficultés de détail, s'est repliée sur elle-même en oubliant le problème principal dont elle était partie.

Le problème principal auquel s'attaquait Boorse dans ses trois articles médicaux⁸ parus entre 1975 et 1977 était pourtant bien celui de la psychiatrie. Il est intéressant à cet égard de rappeler que ces trois articles, publiés dans des revues différentes, sont issus d'un même long manuscrit original, qui fut divisé pour les besoins de la publication scientifique (BOORSE 1997, p. 101). Or dans ce contexte, l'article "What a theory of mental health should be" de 1976, aujourd'hui négligé, aurait dû d'une certaine manière constituer la clef de voûte de la réflexion boorsienne, car c'est dans cet article que l'auteur tente de résoudre les deux difficultés théoriques les plus grandes concernant l'application d'une définition générale du phénomène pathologique à la sphère du mental. Première difficulté: en quel sens peut-on parler de «fonctions naturelles» dans le domaine du mental? Deuxième difficulté: une telle caractérisation de la pathologie mentale est-elle viable sur le plan philosophique? Ne fait-elle pas le jeu d'un dualisme ontologique de l'esprit et du corps?

2.1.1 ➡ «Il ne peut y avoir de santé mentale que s'il y a des fonctions mentales»

Le cœur de l'argumentation de Boorse dans son article de 1976, "What a theory of mental health should be", est de défendre l'idée qu'il est raisonnable de concevoir un certain nombre de fonctions biologiques que l'esprit humain remplirait. Nul besoin pour cela de s'appuyer sur une définition pointilleuse de ce qu'est une fonction biologique. Il

8. "On the Distinction Between Disease and Illness", "Health as a theoretical concept" et "What a theory of mental health should be".

s'agit seulement de parier sur la capacité des sciences psychologiques de parvenir à définir un ensemble de traits psychologiques communs à *Homo sapiens* dont les dysfonctionnements produiraient des syndromes cliniques caractéristiques. Cette possibilité, d'un point de vue purement théorique, présuppose simplement deux choses : d'abord qu'on puisse parler de causalité sur le plan des phénomènes mentaux ; ensuite, qu'on puisse mettre en évidence des relations causales typiques, sur le plan mental, du fonctionnement de l'esprit humain. Ce sont là des présupposés philosophiques lourds, mais qui suffisent si on les accepte pour caractériser de manière générale des «fonctions mentales».

Concernant la question de la causalité entre phénomènes mentaux, Boorse fait valoir l'argumentation de Davidson dans son célèbre article de 1963, "Actions, reasons, and causes" (DAVIDSON 1963) : rien n'interdit de reconnaître à certains événements mentaux la possibilité d'agir comme de véritables causes de l'action. Ce point est déterminant, car il implique que contrairement aux critiques qui lui ont parfois été adressées, l'approche de Boorse n'est pas du tout réductible à une défense ce qu'on appelle communément la «psychiatrie biologique». Boorse entend le concept de «biologie» au sens d'une «science de la vie» suffisamment large pour recouvrir aussi bien la physiologie que la psychologie. En aucun cas pour lui, il ne s'agit de réduire la biologie aux processus «physico-chimiques». En 1997, Boorse réaffirmera avec fermeté qu'il ne voit aucune objection à la possibilité que certaines maladies mentales soient de nature purement psychologique.

S'agissant de la deuxième question, relative aux liaisons causales typiques, Boorse se réfère aux travaux de Freud, de Piaget ou encore de Chomsky : il semble bien que la psychologie soit en mesure de définir certains fonctionnements caractéristiques de l'esprit humain. Toute la difficulté sera de ne pas négliger la plasticité du fonctionnement mental, sans doute plus grande que la plasticité du fonctionnement physiologique du corps. Mais l'idée principale repose sur la conviction qu'il n'est pas excessif de reconnaître une régularité biologique dans certains mécanismes particuliers de l'esprit, dont les dérèglements peuvent être caractéristiques de diverses pathologies. Boorse s'estime en droit de conclure :

Si certains types de processus mentaux exercent des fonctions ordinaires [*standard*] dans le comportement humain, il n'y a aucune raison de ne pas caractériser les entraves non naturelles à ces fonctions des

maladies mentales [*mental diseases*], exactement comme dans le cas de la physiologie. (BOORSE 1976, p. 64.)⁹

L'auteur reste néanmoins très évasif sur la réalité et le nombre de ces fonctions mentales. Ce qui compte avant tout ici, sur le plan théorique, c'est de faire valoir la pertinence d'une telle approche. En ce sens, il n'est pas sans intérêt de rappeler que la théorie psychanalytique de Freud apparaît à ses yeux en 1976 comme celle qui offre le meilleur modèle de ce qu'une «théorie de la santé mentale *devrait être*» (*ibid.*, p. 78, nous soulignons)¹⁰. Cette anecdote indique au moins une chose : l'enjeu est méthodologique avant d'être une question d'orientation théorique pour la psychiatrie.

2.1.2 ➔ La psychiatrie, entre corps et esprit

Mais une telle approche méthodologique, opposant les maladies du corps aux maladies de l'esprit, ne revient-elle pas à enfermer la psychiatrie dans le vieux dualisme cartésien ? Le deuxième point qu'entend démontrer Boorse, c'est que la psychiatrie en tant que discipline médicale n'est pas condamnée au dualisme, pas plus qu'elle n'est condamnée à n'être qu'une neurologie ratée. S'appuyant sur les travaux de la philosophie de l'esprit des années 1970, et notamment Davidson, Putnam, ou Fodor, il cherche à montrer qu'il est possible d'accorder une certaine autonomie au mental sans forcément être obligé de se réclamer du dualisme cartésien. Par exemple, la théorie de l'identité de l'esprit (*Identity Theory*), dont Boorse se contente de résumer les grandes lignes, permet d'envisager que tous les événements mentaux soient des événements physiques, sans pour autant que chaque événement mental particulier soit strictement corrélé à un événement physique particulier. Cette forme de matérialisme raffiné concilie ainsi le fait que «désirer manger un homard à l'instant t» est un événement mental qui a nécessairement une inscription cérébrale particulière, mais qui n'est pas forcément susceptible d'être identifié à un type d'événement neural particulier.

9. Sauf indication, toutes les citations qui suivent des textes anglais sont traduites par nous.

10. Boorse se justifie ainsi : «Formellement parlant, la théorie psychanalytique est la meilleure conception de la santé mentale que nous ayons. Elle suit de près le modèle physiologique en postulant trois sous-structures mentales, le ça, le moi et le surmoi, et elle assigne à chacune une fonction définie.»

Autrement dit, tout en se revendiquant d'une attitude matérialiste, il est possible de défendre le point de vue d'une dualité épistémique des sciences du mental et du cérébral. À l'époque où il écrit son article, les études relatives à cette approche philosophique qui peut être qualifiée de «fonctionnaliste» sont florissantes. Boorse cherche simplement à souligner l'importance que ces considérations nouvelles issues de la philosophie de l'esprit représentent pour la psychiatrie : «Mon but est [...] de montrer en quoi il n'est pas évident du tout que les psychiatres qui rejettent le dualisme cartésien détruisent nécessairement l'autonomie de leur discipline» (*ibid.*, p. 66).

De fait, le fonctionnalisme en philosophie de l'esprit permet de répondre à l'une des objections fortes de Szasz à l'encontre de la psychiatrie : on est en droit de parler de «maladies mentales» quand bien même aucun dysfonctionnement physiologique du cerveau n'est observable. La mise en évidence d'une lésion caractéristique du cerveau n'est plus un élément à charge de preuve pour que l'on puisse parler légitimement de maladie mentale. La psychiatrie peut se réclamer d'un espace d'investigation qui lui est propre, indépendant de la neurologie, et qui l'autorise pleinement à se prévaloir du vocabulaire médical.

Le projet théorique de la psychiatrie a donc un sens, et ni le dualisme cartésien, ni le matérialisme ne sont des options métaphysiques déterminantes susceptibles de mettre en cause l'autonomie de ce projet. La seule question véritablement déterminante, c'est la question de savoir si on est en droit ou non de parler de «causalité» pour les phénomènes mentaux et de «fonction mentale». Si l'on s'accorde sur ces deux points, alors une science du fonctionnement normal et anormal de l'esprit humain est envisageable, même si elle reste à construire. À ce niveau de généralité, on peut donc dire que Boorse a tenu son pari, en dépit des nombreuses difficultés théoriques qui demeurent présentes¹¹.

Mais quel est l'enjeu précis pour la psychiatrie? Fondamentalement, c'est celui de défendre l'idée que cette science du fonctionnement normal et anormal de l'esprit humain si vivement souhaitée soit amenée à four-

11. Comme cette difficulté particulière, soulevée par Fulford, que l'analyse que propose Boorse du concept de fonction en termes de contribution d'un sous-système au système entier ne semble pas toujours pertinente en psychiatrie, où la plupart des troubles semblent mettre en cause des fonctions centrales.

nir l'unique critère pertinent de démarcation du normal et du pathologique. La clinique à elle seule n'est jamais suffisante comme marqueur du phénomène pathologique. Elle n'en donne au mieux qu'un indice, et Boorse est formel sur l'idée que dans la détermination du normal et du pathologique, il doit revenir à la science et à la médecine théorique de surplomber la clinique¹². Dans cette perspective, Boorse n'hésiterait pas, semble-t-il, à reprendre à son compte le mot de Leriche: «Si on veut définir la maladie, il faut la déshumaniser» (cité par CANGUILHEM 1998, p. 53). Seulement voilà: comment faire? Que peut-on faire en psychiatrie tant qu'on ne dispose pas positivement d'une science du fonctionnement normal de l'esprit humain? Et quand bien même on en aurait une certaine idée, comment procéder pour distinguer les marges du fonctionnement normal du dysfonctionnement sur le plan mental?

2.2 ⇨ Le problème de la démarcation: l'approche bio-statistique

On peut donc porter au crédit de Christopher Boorse d'avoir su redonner une certaine pertinence au vieux projet d'une psychiatrie scientifique susceptible de caractériser objectivement les dérèglements mentaux. Le problème qu'il reste à résoudre est néanmoins celui de pouvoir traduire cette légitimité théorique dans une pratique empirique, autrement dit, de fournir les moyens théoriques permettant de distinguer objectivement en psychiatrie les marges du normal et du pathologique (2). Boorse aborde ce problème sous un angle général dans l'article de 1977, "Health as a theoretical concept", mais c'est dans la synthèse qu'il propose de ses travaux en 1997 qu'on trouve les éléments de réponse les plus déterminants sur cette question. Amené à affiner sa position théorique, il accepte alors l'étiquette de «théorie bio-statistique» dont le philosophe suédois Lennart Nordenfelt s'était servi pour mettre l'accent sur les deux éléments clefs de sa conception, à savoir les idées de fonction biologique et de normalité statistique. Toute l'importance du rôle que ces deux idées jouent dans la BST se trouve résumée dans cette définition de la santé et de la maladie que Boorse considère comme sa proposition définitive:

12. Cf. BOORSE (1997, p. 48): "But one cannot go in the opposite direction, defining disease from clinical concepts."

1. La *classe de référence* est une classe naturelle d'organismes partageant un *design* fonctionnel uniforme; en particulier, un groupe de même âge et de même sexe dans une même espèce.
2. La *fonction normale* d'une partie ou d'un processus, chez les membres de la classe de référence, est la contribution statistiquement typique qu'elle ou qu'il fait à la survie et à la reproduction individuelles.
3. Une *maladie* est un type d'état interne qui est soit une altération d'une capacité fonctionnelle normale, c'est-à-dire une réduction d'une ou de plusieurs capacités fonctionnelles en dessous du niveau d'efficacité typique, soit une limitation de la capacité fonctionnelle causée par des agents environnementaux.
4. La *santé* est l'absence de maladie. (BOORSE 1997, p. 7-8.)

La mention centrale ici est celle d'une «typicité statistique» des contributions fonctionnelles: Boorse veut se donner les moyens de traduire son approche naturaliste de la norme biologique dans le langage de la norme statistique. En ce sens, on peut dire que l'originalité de son approche bio-statistique, comme l'a bien résumée Élodie Giroux dans un récent article (GIROUX 2009), est d'une certaine manière de vouloir concilier la physiologie de Claude Bernard avec la conception de l'homme moyen d'Adolphe Quételet. Car le problème pour lui est le suivant: on ne peut pas parler de «fonction normale» d'un organe ou d'un processus sur le plan biologique sans avancer, de fait, une certaine idée du «*design* de l'espèce» (*species design*), c'est-à-dire de l'ensemble des caractéristiques plutôt uniformes que partagent les membres d'une même espèce. Or dès 1977, Boorse s'était montré sensible au danger qu'il pouvait y avoir de verser de la sorte dans une conception essentialiste des propriétés fonctionnelles partagées par une même espèce, le risque étant de négliger les phénomènes de variabilité biologique et de réifier le *design* de l'espèce. Aussi, la définition qu'il propose de la fonction normale comme d'une «contribution statistiquement typique» d'un trait à la survie ou la reproduction de l'individu a pour but de répondre aux accusations d'essentialisme qui pourraient lui être faites, toute en donnant les gages de son caractère factuel à l'intérieur d'une classe de référence donnée. Boorse reprend sur ce point l'opposition devenue classique que proposait Ernst Mayr entre le type de pensée typologique et le type de pensée populationnel en biologie, et il se défend en affirmant que la «la BST est indiscutablement populationnelle» (BOORSE 1997, p. 39): loin de sacrifier à un quelconque essentialisme (ou pensée typologique), sa théorie met en valeur le fait que le *design* de l'espèce n'est rien d'autre qu'une «abstraction statistique» précaire (BOORSE 1977, p. 558). Or il en

va de même selon lui, mais à une autre échelle, du phénomène pathologique. Dans "A Rebuttal on Health", Boorse accompagne sa définition du schéma reproduit ci-après (BOORSE 1997, p. 8).

L'ensemble de ce qui est engagé théoriquement dans cette simple représentation graphique mériterait un ouvrage. Contentons-nous de constater que la courbe représentée, qui semble proposer le moyen d'une démarcation quantitative entre les phénomènes normaux et les phénomènes pathologiques pour une fonction donnée, ne fait en réalité qu'illustrer théoriquement les enjeux de l'approche populationnelle : on peut parler de phénomène pathologique lorsque l'efficacité qu'on prête à une fonction est si faible qu'elle ne rend plus compte de sa *fitness* au sein d'une population. L'intérêt de cette courbe s'arrête là : concilier l'idée de variation naturelle avec celle d'une objectivité du phénomène pathologique. Mais il serait trompeur de croire qu'elle puisse prétendre avoir une valeur heuristique ou ouvrir sur certaines pistes d'investigation empirique en épidémiologie médicale.

Un premier malentendu, sur le plan médical et psychiatrique, serait de voir dans cette courbe une caractérisation du phénomène pathologique comme simple déviation statistique d'un trait. Boorse s'est toujours clairement opposé à cette vieille idée¹³. Ce qui est anormal (au sens statistique) n'est pas forcément pathologique. La rareté d'un trait ne suffit pas à attester son caractère pathologique, et c'est là un point crucial du débat antipsychiatrique autour de ce qu'il est convenu d'appeler la normalisation des conduites. L'ingéniosité de l'approche de Boorse est de se concentrer sur la répartition statistique de l'*efficacité d'une fonction*, non de la *présence d'un trait*. La nuance importe, car elle indique que la courbe, contrairement aux apparences, est épaisse. Elle est épaisse de tout un passé évolutif¹⁴. Elle n'a en effet de sens que comprise comme la coupe synchronique (approche statistique) d'un processus diachronique général qui est présupposé : le désavantage biologique, soit la baisse de la reproduction et de la survie individuelles au sein de la population lorsque la fonction en jeu n'est plus efficace. Autrement dit, la référence

13. Cf. BOORSE (1977, p. 546) : la normalité statistique n'est une condition ni nécessaire ni suffisante de la santé. Pour une discussion attentive de cet aspect, nous renvoyons à ANANTH (2008, p. 18 et suiv.).

14. Cf. BOORSE (1997, p. 66) : "So for me some of the past affects what is species-typical."

au rôle que la sélection naturelle exerce dans l'histoire évolutive des organismes permet à Boorse d'expliquer la normalité du fréquent, à l'instar de ces «causes constantes» auxquelles faisait référence Adolphe Quételet, mais bien plus subtilement que ce dernier – lequel ne parvenait à rendre raison de la normalité du fréquent qu'en la faisant découler d'un dessein divin.

Mais il y a un deuxième malentendu à dissoudre, en quelque sorte l'inverse du précédent. Ce serait de négliger la question de l'efficacité fonctionnelle en se concentrant uniquement sur la mise en évidence d'un désavantage biologique. Cet autre malentendu tenace, que Boorse entend rectifier¹⁵, a lui aussi des antécédents en psychiatrie, puisque des tentatives ont été faites de corrélérer la présence de certains syndromes cliniques avec des critères comme la fertilité ou l'espérance de vie dans l'idée que cela fournirait une preuve empirique de leur caractère pathologique (cf. KENDELL 1975).

En croisant les approches biologiques et statistiques, Boorse répond de manière convaincante à ces deux malentendus dont il perçoit le caractère pernicieux en psychiatrie. Mais c'est au prix d'une «idéation statistique» qui, pour neutre qu'elle soit sur le plan axiologique, ne peut avoir aucune valeur pragmatique en médecine, et en particulier en psychiatrie. Élodie Giroux, dans l'article précédemment cité, a insisté sur les difficultés qu'il y aurait par exemple à tirer de la conception théorique de Boorse des conclusions intéressantes pour l'épidémiologie. On peut sommairement évoquer trois difficultés empiriques qui rendent cette théorie tout à fait impraticable pour fixer le critère du normal et du pathologique. Tout d'abord, la courbe en cloche censée représenter l'efficacité d'une fonction particulière au sein de la population est un schéma idéal qui fait explicitement abstraction des facteurs environnementaux et sociaux. Or une telle abstraction est trompeuse, et même illusoire : il n'y a aucune raison *a priori*, pour reprendre un argument que Maurice Halbwachs faisait déjà valoir en son temps contre Quételet, que les mesures empiriques de n'importe quel fait biologique ou social dans une population donnée reproduise la forme exacte d'une courbe de Gauss. Le deuxième point, c'est que rien

15. Boorse reproche ainsi à Wakefield de l'avoir rangé parmi les partisans du désavantage biologique, aux côtés de Scadding ou de Kendell. Cf. BOORSE (1997, p. 65).

ne justifie non plus le seuil arbitraire des 2,5 %¹⁶ retenus pour définir le phénomène pathologique. Boorse est certes conscient de l'arbitraire de ce seuil, et il laisse – une fois n'est pas coutume – à la clinique le soin d'en décider. Mais il reste que ce chiffre, qui cherche à conférer un sens biologique à la variation normale statistique, n'a strictement aucune valeur pragmatique, spécialement en épidémiologie où de nombreux facteurs biologiques, génétiques, sociaux et environnementaux sont nécessairement mis en balance avec la gravité des troubles pour ajuster le seuil qu'on considère comme pathologique, et qui peut être haut pour certaines populations¹⁷, ou très bas pour d'autres¹⁸. Le seuil présenté par Boorse des 2,5 % est non seulement arbitraire, mais il est inopérant et dépourvu de sens épidémiologique. Le dernier point, qui aggrave les insuffisances précédentes, c'est que toute mesure de l'efficacité d'une fonction ne s'effectue bien souvent, Boorse le reconnaît lui-même (BOORSE 1997, p. 50), qu'imparfaitement et indirectement à travers la mesure d'indicateurs biologiques (voire de tests psychométriques). L'ensemble de ces trois critiques – c'est le point important

16. 95 % de ce qu'on appelle la variation normale d'une distribution gaussienne est comprise dans un intervalle incluant deux écarts types de part et d'autre de la moyenne. Comme Boorse ne s'intéresse qu'à la frange basse de la courbe d'efficacité, il est conduit à fixer le seuil du phénomène pathologique autour de 2,5 %.

17. La prévalence en France de la malvoyance est estimée à 6,8 % pour les personnes de 70 ans, et à 14 % pour les personnes de 80 ans (données INSEE, 1987). Bien sûr, on pourrait considérer, en suivant Boorse, qu'à l'intérieur de chaque population de classe d'âge, seuls les 2,5 % les plus malvoyants sont vraiment pathologiques, le reste étant laissé à la médecine de confort. Pourtant, quand on sait la sévérité du trouble (est considérée comme malvoyante une personne dont l'acuité visuelle du meilleur œil, après correction optique, est inférieure à 4/10), il peut sembler étrange de parler encore de «médecine de confort». Remarquons au passage que l'approche essentialiste de Wakefield, que nous allons détailler ensuite, aboutirait sur cet exemple précis à une conclusion plus simple et consensuelle, mais pas forcément mieux fondée en raison. L'idée serait que la fonction naturelle des yeux, même pour une vieille personne, est de voir, et que cette seule considération suffit pour considérer le problème de la malvoyance, quel que soit l'âge, comme un problème pathologique.

18. Bien souvent dans ces cas-là, le critère le plus parlant n'est ni le sexe, ni la classe d'âge de la population, mais le pays où l'enquête est menée.

sur lequel nous voulons insister – rend le critère biologique de démarcation entre le normal et le pathologique tout à fait impossible à fixer objectivement, ne serait-ce que de manière approximative et indicative, quand on se tourne du côté de la psychiatrie. Ce n'est pas un hasard, nous semble-t-il, si l'on ne trouve nulle part de discussion approfondie sur une possible application pratique de l'approche bio-statistique de Boorse au domaine mental¹⁹. La raison, encore une fois, en est simple : on voit mal comment un débat aussi difficile à démêler concernant des maladies physiques bien connues (comme la tuberculose ou l'anémie falciforme) pourrait être productif sur le terrain des maladies mentales, même parmi celles qui arrachent un consensus raisonnable (comme la schizophrénie ou la dépression). De fait, dans le domaine des maladies mentales, la difficulté n'est pas seulement de séparer le poids relatif des facteurs biologiques et culturels. Elle est aussi d'identifier clairement les fonctions mentales qui sont supposées défaillantes, et d'être capable d'en mesurer l'efficacité au sein d'une population. À supposer par exemple, pour reprendre une hypothèse développée par HORWITZ & WAKEFIELD (2007), que la dépression soit un dysfonctionnement pathologique de la tristesse, l'idée qu'on puisse un jour obtenir une mesure ne serait-ce qu'approximative de l'efficacité naturelle d'une telle fonction biologique au sein d'une population donnée est tout bonnement irréaliste.

Il faut donc bien conclure que si Boorse propose une théorie stimulante qui répond – même si c'est de manière indirecte – au défi théorique lancé par Canguilhem de traduire quantitativement la différence entre les phénomènes normaux et pathologiques, la solution qu'il apporte n'a aucune chance de pouvoir clarifier le débat psychiatrique dans sa réalité effective. Boorse est conscient, bien évidemment, des limites pragmatiques de sa théorie, et il entend laisser le soin à la médecine de définir et de délimiter les maladies «suivant les bases scientifiques les plus commodes» (BOORSE 1997, p. 64). Mais l'approche bio-statistique

19. Les thèses centrées sur la question psychiatrique de Rachel Cooper, *Classifying Madness: A philosophical examination of the Diagnostic and Statistical Manual of Mental Disorders* (PhD Thesis, Cambridge, 2005) et de Jonathan Tsou, *The Reality and Classification of Mental Disorders: Issues in the Philosophy of Psychiatry* (PhD Thesis, Chicago University, 2008), bien qu'elles discutent attentivement la proposition de Boorse, se contentent pour l'essentiel de reprendre les objections biologiques générales.

qu'il préconise est vouée, nous semble-t-il, à achopper sur des difficultés en psychiatrie qui ne sont pas seulement conjoncturelles. En mettant entre parenthèses la clinique psychiatrique pour en fonder la science, Boorse a oublié que la psychiatrie n'avait encore très précisément que sa clinique pour légitimer sa pratique.

3 ⇨ L'approche essentialiste de Wakefield

3.1 ⇨ L'analyse du dysfonctionnement préjudiciable

Jerome Wakefield, contrairement à Christopher Boorse, est un spécialiste des questions psychiatriques. Son article séminal, "The Concept of Mental Disorder: On The Boundary Between Biological Facts and Social Values", initialement paru dans la revue *American Psychologist* en 1992, est l'un des articles les plus cités et les plus commentés dans le débat épistémologique en psychiatrie de ces vingt dernières années. Au rebours du naturalisme impénitent de Boorse²⁰, Wakefield défend un «normativisme faible». Dans le domaine de la psychiatrie, cela revient à reconnaître l'irréductibilité de l'immixtion des valeurs sociales dans la constitution des catégories diagnostiques, mais à contrebalancer cet aspect fâcheux par la mise en évidence d'une composante factuelle dysfonctionnelle légitimant l'appartenance pleine et entière des troubles mentaux au modèle médical. Pourtant, malgré les divergences d'étiquette, sa position générale sur le plan théorique est très proche de celle de Boorse. La différence tient surtout au fait que Wakefield cherche un outil pratique, applicable à la question psychiatrique, qui a pour prétention de servir de test ou de filtre permettant d'établir la légitimité ou non des critères diagnostiques des troubles mentaux retenus dans les classifications d'usage, là où Boorse était davantage préoccupé à faire valoir une position théorique marginale encore à son époque. De fait, lorsque Wakefield ajoute la clause de préjudiciabilité (*harmful*) à son analyse, c'est davantage pour faire valoir l'idée, possible en théorie – et avec laquelle Boorse finalement s'accordait –, qu'il peut y avoir des dysfonctionnements qui ne méritent pas d'être considérés comme pathologiques sur le plan clinique pour la simple raison qu'ils ne sont pas préjudiciables et peuvent même parfois être profitables à un individu. Autrement dit, Wakefield semble poursuivre l'effort de Boorse en orien-

20. *Ibid.*, p. 5: "I am an unrepentant naturalist."

tant le programme de l'analyse conceptuelle sur un plan explicitement pratique et clinique. D'ailleurs, Boorse, dans son article rétrospectif de 1997, n'hésitera pas à voir dans la HDA de Wakefield davantage une prolongation de l'effort de la BST qu'une théorie rivale : «L'approche de Wakefield, comme je l'ai dit, est en substance celle de la BST, sauf qu'il remplace ma théorie de la fonction par celle de Wright, et qu'il ajoute une clause de préjudiciabilité pour rendre l'idée de dysfonctionnement importante sur le plan clinique» (*ibid.*, p. 66).

La parenté entre les deux théories est indéniable. Pourtant, les différences sont plus grandes qu'elles ne laisseraient apparaître de prime abord. Globalement, la clause (1) du contrat objectiviste est la même : on est en mesure de parler du normal comme d'un fait pour autant qu'on se place sur le plan du fonctionnement biologique. Mais la solution à la clause (2) est assez différente. Contrairement à Boorse, Wakefield semble refuser le défi lancé par Canguilhem d'avoir à traduire d'une manière quantitative la différence entre le normal et le pathologique. Hostile à toute caractérisation statistique du phénomène pathologique, Wakefield opte plus ou moins explicitement pour une position essentialiste, simple en apparence, mais dont la justification théorique, qui s'appuie sur la théorie de l'évolution, est audacieuse.

Sur quoi repose le critère de démarcation entre le normal et le pathologique chez Wakefield ? Il repose, pourrait-on dire, sur une différence d'essence ou de nature qui, si l'on tient attentivement compte des pré-supposés de la HDA, ne peut être traduite de manière univoque par aucune enquête épidémiologique ni même être formalisée quantitativement d'aucune manière théorique. La différence entre une dépression et une tristesse profonde, par exemple, est qualitative plutôt que quantitative. Bien qu'étant objective et fondée biologiquement, cette différence ne peut s'apprécier que de manière contextuelle et c'est pourquoi elle exige selon Wakefield un certain tact clinique. Chez le dépressif, la tristesse est dysfonctionnelle, au sens où il est évident d'un point de vue clinique qu'elle ne remplit plus la fonction biologique qu'elle est censée remplir chez l'individu normal. En revanche, chez celui qui vient de perdre un être cher – pour prendre une situation caractéristique –, la tristesse est normale (même si elle peut être intense) : elle répond à un événement de perte qui, aussi indéterminée que soit notre connaissance du mécanisme en question, peut être considéré comme l'élément déclencheur du type de réponse normale pour laquelle l'émotion de tristesse

a été naturellement façonnée par l'évolution. Nous ne nous arrêterons pas sur les difficultés considérables que cette proposition soulève dans le cas précis de la dépression. Notre objection est plus fondamentale et s'adresse à la capacité que Wakefield prête à sa théorie de pouvoir éclairer la plupart des controverses nosologiques en s'appuyant sur ce genre de raisonnement. Le point précis de notre critique porte sur la conception théorique particulière de la fonction biologique sur laquelle Wakefield fait reposer sa définition du normal et du pathologique. En voulant faire d'une définition théorique (plus audacieuse qu'il ne voudrait l'avouer) un instrument opératoire susceptible de régler la pratique clinique, la HDA de Wakefield prête davantage le flanc à la critique que la théorie de Boorse. Pour que sa définition soit utile à la psychiatrie, Wakefield est en effet obligé de négliger certaines subtilités théoriques que Boorse a pris soin d'explicitier. D'abord, tandis que Boorse insiste sur le fossé qui sépare la pathologie comme science théorique de la clinique comme pratique médicale²¹ – ménageant ainsi un espace à « plusieurs niveaux » (*ibid.*, p. 101) plutôt conciliant et dont l'articulation n'est pas fixée une fois pour toutes –, Wakefield propose une analyse qui prétend embrasser dans un même geste les intuitions cliniques les plus traditionnelles et les hypothèses évolutionnistes les plus récentes. Là où Boorse manifestait en outre une prudence minimale à l'égard des usages profanes du concept de maladie, Wakefield n'hésite pas à faire appel au sens commun à différentes étapes de la justification conceptuelle de son approche. L'analyse de Wakefield se nourrit comme celle de Boorse des intuitions du sens commun (cf. WAKEFIELD 2002)²², mais elle s'engage bien davantage sur un certain plan théorique, puisqu'elle repose explicitement sur une conception évolutionniste des fonctions biologiques. Cet écart, selon nous, soulève une difficulté particulière dans la théorie de Wakefield qu'on ne trouve pas chez Boorse, du moins pas d'une manière aussi aiguë. Ce dernier en effet avait pris la peine de faire remarquer que sa théorie bio-statistique ne reposait sur aucune approche définitive de la fonction biologique et que son travail sur le

21. *Ibid.*, p. 11 : "It aims at a pathologist's concept of disease, not a clinician's [...]."

22. L'auteur écrit : "I assume that the concept of disorder is shared by lay people and professionals."

concept de santé pouvait être séparé de celui qu'il avait entrepris sur le concept de fonction biologique (cf. BOORSE 1997, p. 10)²³.

La position de Wakefield est en revanche beaucoup plus délicate à justifier. Même s'il insiste, lui aussi, sur le fait que son analyse est séparable de la conception étiologique de la fonction qu'il privilégie, il est à noter que toute la valeur pratique et clinique qu'il attribue à la HDA repose néanmoins sur elle. Enlevez l'ancrage évolutionniste de la HDA de Wakefield, vous perdez tout son sel. Mais le problème est plus grave : Wakefield doit justifier l'unité du concept de fonction qu'il mobilise. De fait, l'idée qu'une fonction soit un effet sélectionné, suivant la théorie étiologique, n'est pas du tout une idée intuitive. Et comme l'a bien montré Kenneth SCHAFFNER (1993)²⁴, elle est loin de coïncider avec l'idée en usage dans les sciences biomédicales. Or Wakefield se doit de résorber ce hiatus apparent.

3.2 ➔ La «boîte noire essentialiste»

Bien qu'elle constitue une référence ancienne et prédominante dans le travail de Wakefield²⁵, la «boîte noire essentialiste²⁶» ne commence

23. Certes, on pourrait reprocher à Boorse, à travers ses articles, la tendance à défendre sa théorie bio-statistique en s'appuyant volontiers sur la conception de la fonction qu'il a toujours défendue comme étant la bonne, une conception «orientée vers les buts» (*goal-directed*). Mais même si c'est la cas, il est à noter que la conception qu'il privilégie de la fonction comme «contribution actuelle à un but» est suffisamment souple et large pour intégrer la plupart des acceptions particulières du terme de fonction aussi bien en biologie que dans la pratique médicale, et cela tout en s'accordant avec l'usage commun.

24. Schaffner souligne (p. 389-390) que les explications fonctionnelles évolutionnistes, pour les sciences biomédicales, sont nécessaires mais «si empiriquement fragiles qu'on peut presque les dire métaphysiques».

25. Il est clair que la référence à la «boîte noire essentialiste» est ancienne dans le travail de Wakefield (on en trouve la référence dès 1991 dans un article traitant de la question des émotions chez Freud), et qu'elle s'inspire largement, comme il le reconnaît lui-même, des travaux de Putnam. Concernant le problème précis qui nous occupe, la première référence digne d'intérêt date néanmoins de 1997.

26. Nous proposons ainsi de traduire "*black box essentialism*", souvent rendu par Wakefield aussi par "*black box essentialist account of concepts*". Il ne

à être mobilisée que tardivement pour justifier la HDA, et cela explicitement en réponse à certaines objections qui lui sont adressées. Nous pensons que cette conception, qui se présente comme valant de manière générale pour tous les concepts scientifiques, constitue à la fois le cœur et le talon d'Achille de l'analyse proposée par Wakefield du trouble mental.

Clairement, la référence à cette conception apparaît à un moment où Wakefield doit répondre à une double exigence pour défendre sa HDA: (a) il doit assurer l'unité du concept de fonction, que ce soit dans le champ psychiatrique, médical ou biologique en général; (b) il doit proposer une méthode qui permette de circonscrire un ensemble de dysfonctionnements mentaux, et ce d'une manière qui puisse apparaître comme légitime malgré la précarité persistante de nos connaissances scientifiques en psychiatrie.

3.2.1 ➔ L'unité du concept de fonction

Le travail initial de Wakefield s'inscrit dans le cadre d'une analyse conceptuelle dont il rappelle le principe :

Dans le cadre d'une analyse conceptuelle, les versions proposées d'un concept sont confrontées aux jugements assez peu controversés et largement partagés concernant ce qui tombe ou ne tombe pas sous le concept. Pour autant que l'analyse parvienne à expliquer ces jugements non controversés, elle est considérée comme confirmée, et une analyse suffisamment confirmée peut alors être utilisée comme un guide pour réfléchir sur les cas plus controversés. (WAKEFIELD 1992a, p. 233.)

S'il précise que les jugements consensuels ne sont pas forcément supposés corrects, une bonne analyse doit néanmoins permettre de distinguer les jugements corrects de ceux qui ne le sont pas. L'enjeu est de parvenir par cette méthode à fournir une *définition descriptive* du trouble mental, qui soit adéquate à l'usage ordinaire du concept chez les médecins et même dans la perception commune. Tout écart avec la signification ordinaire apparaît comme un problème gênant pour la définition dont il faut pouvoir rendre compte. L'ambition fondamentale

nous semble pas que nous dénaturions le sens de cette étiquette en faisant de l'essentialisme un attribut de la boîte noire plutôt que l'inverse. Comme la définition qu'en propose Wakefield le montre bien, "black box" est une modalité intrinsèque du type d'essentialisme qu'il s'agit de défendre.

de la HDA est bien de *décrire* et non de *réformer* notre conception du phénomène pathologique.

Pourtant, dès son article de 1992, Wakefield fonde sa démarche sur la psychologie évolutionniste, en présupposant deux choses : d'abord, que seul le concept évolutionniste de fonction est susceptible de donner un fondement scientifique à la qualification de dysfonctionnement ; mais que néanmoins, il y a une unité profonde du concept de fonction, tel qu'il est entendu intuitivement, tel qu'il est mobilisé dans les sciences médicales, et enfin tel qu'il est compris à la lumière de la théorie de l'évolution. Aussi, au détour d'une discussion avec Christopher Megone au sujet du sens de la notion de fonction chez Aristote (WAKEFIELD 2000), Wakefield est amené à préciser sa position par rapport à un problème important soulevé par Ruth Millikan et Karen Neander, celui d'une éventuelle difficulté qu'il y aurait à concilier les présupposés de l'analyse conceptuelle avec une conception étiologique de la fonction. En mettant au centre de son analyse conceptuelle du phénomène pathologique une conception particulière de la fonction naturelle en biologie, Wakefield est curieusement conduit – davantage que Boorse, spécialiste de philosophie de la biologie mais qui s'est montré plus prudent dans sa proposition de philosophie médicale –, à se déplacer du terrain de la médecine sur celui de la philosophie de la biologie pour prendre parti dans le débat extrêmement pointu ouvert par les articles de Wright et de Cummins dans les années 1970. Il doit notamment répondre à une critique très pertinente de la philosophe américaine Ruth Millikan qui considérait qu'un hiatus irréductible existait, pour certains termes comme celui de « fonction », entre les caractérisations issues du sens commun et les caractérisations fondées théoriquement. Assumant pour sa part le caractère réformiste de sa conception théorique des fonctions biologiques, qu'elle refusait d'inscrire de force dans le prolongement de nos intuitions communes, Millikan avait au passage des mots durs à l'encontre de la méthode de l'analyse conceptuelle en philosophie, si communément convoquée pour justifier le bien-fondé des concepts communs dans la tradition analytique :

Je suis maintenant fermement convaincue que « l'analyse conceptuelle », comprise comme la recherche des conditions nécessaires et suffisantes de l'application des termes, ou comme la recherche de critères d'application par référence à quoi un terme a la signification qu'il a, est un programme confus, une chimère philosophique, la mise en carré

d'un cercle, l'enfant mal conçu d'une vue erronée de la nature de la pensée et du langage. (MILLIKAN 1989, p. 290.)

Sur cette question particulière, la philosophe Karen Neander a répondu à Millikan en défendant les bienfaits de l'analyse conceptuelle et en montrant que celle-ci était compatible et même complémentaire avec la recherche d'une définition théorique de certains termes scientifiques (NEANDER 1991). Wakefield, pour sa part, dont toute la démarche repose sur l'analyse conceptuelle, renvoie avec assurance ces deux auteurs dos à dos, soutenant que le concept de fonction n'est pas *directement* lié à l'idée de sélection naturelle, ni en vertu d'une analyse conceptuelle, ni en vertu d'une définition théorique quelconque. «Que les fonctions existent parce qu'il y a la sélection naturelle, c'est là une découverte scientifique, pas une vérité conceptuelle» (WAKEFIELD 2000, p. 39), affirme-t-il. Et il justifie cela par un processus en deux étapes : il y aurait d'abord un sens commun au concept de fonction, intuitivement partagé par tous, de tout temps (et donc aussi bien par Aristote, par Harvey que par nous). C'est ce sens commun que mettrait en évidence l'analyse conceptuelle. Puis il y aurait une «découverte moderne», selon laquelle c'est la sélection naturelle qui explique l'«essence» des processus fonctionnels (*ibid.*). Pour assurer la continuité entre les deux étapes, Wakefield fait alors référence à ce qu'il appelle la «boîte noire essentialiste», qu'il revendique à la suite des travaux de Putnam, Kripke et Searle sur les *natural kinds*. Voici la définition qu'il en propose :

C'est une approche «essentialiste» au sens où le critère d'adhésion ne consiste pas dans les propriétés qui ont initialement inspiré quelqu'un pour définir le genre, mais dans une certaine propriété «essentielle» sous-jacente qui explique les propriétés de surface observées. Mais cet essentialisme est une «boîte noire» car, plutôt que d'affirmer que les concepts scientifiques sont définis par des essences spécifiques (par exemple, «l'eau est H₂O»), cette approche soutient qu'une telle définition demeure agnostique quant à l'identité de cette essence qu'elle laisse à la science la charge de découvrir. (*Ibid.*, p. 36.)

Cette référence à la «boîte noire essentialiste» est centrale et nous y reviendrons. Mais demandons-nous d'abord ce que recouvre pour la psychiatrie ce questionnement très théorique autour de l'unité du concept de fonction. Il faut bien voir que tacitement, l'enjeu est de défendre *la continuité de la tradition psychiatrique*. Aristote partageait le même concept de fonction que nous, même s'il en ignorait l'explication moderne fournie par la théorie de l'évolution. De même, les psychiatres du XIX^e

siècle partageaient la même idée de la maladie mentale que nous, même s'ils étaient incapables de la justifier scientifiquement. Wakefield refuse le dilemme d'avoir à choisir entre deux paradigmes psychiatriques. La «psychiatrie darwinienne²⁷» qu'il entend promouvoir, malgré la jeunesse de son projet, n'est pas selon lui en rupture avec le discours psychiatrique traditionnel. Bien au contraire, elle en formerait le prolongement nécessaire. La HDA n'impose aucune rupture ni réforme de fond au discours psychiatrique : elle entend bien plutôt le conforter.

3.2.2 ➔ La valeur heuristique de la HDA

Cet aspect est particulièrement saillant dès qu'on se penche sur les vertus heuristiques que Wakefield prête à son approche, et qu'il justifie de nouveau en se référant à la théorie de la «boîte noire essentialiste». De fait, sans une méthode qui permette de circonscrire légitimement le domaine des troubles mentaux en l'absence de connaissances assurées sur le plan scientifique, la définition proposée par Wakefield serait dépourvue de toute utilité en psychiatrie. À quoi servirait en effet de définir comme légitime un trouble qui repose sur un dysfonctionnement préjudiciable, si rien, dans l'état actuel de nos connaissances, ne nous permettait légitimement d'inférer l'existence de certains dysfonctionnements préjudiciables ? La définition serait tout juste utile à titre d'idéal régulateur. Mais elle ne serait certainement d'aucune aide pour résoudre les problèmes actuels auxquels est confrontée la psychiatrie. Or le coup de force de Wakefield, c'est de se prévaloir de son approche évolutionniste pour légitimer la plupart (sinon la totalité) des étiquettes

27. Wakefield fait d'ailleurs très peu usage de cette étiquette popularisée par McGuire et Troisi à la suite de Nesse et Williams, pas plus du reste qu'il n'en accepte les postulats théoriques. La raison en est simple mais profonde : comme l'a très bien mis en valeur MURPHY (2005), l'approche darwinienne des troubles mentaux intègre trois types d'explications possibles : le dysfonctionnement interne (*breakdown*), le fait pour un organisme de ne plus être adapté à un environnement qui a rapidement changé (*mismatch*), et la possibilité paradoxale pour un trouble mental d'être un avantage adaptatif (*persistence*). Or, la HDA de Wakefield n'est compatible qu'avec le premier type d'explication (*breakdown*), ce qui la met en porte-à-faux avec de nombreuses hypothèses évolutionnistes. Il n'en demeure pas moins que l'étiquette générale lui correspond.

de la classification de référence en psychiatrie, le DSM. Cette longue citation mérite une attention particulière :

Si l'on regarde la liste des Troubles recensés par le DSM, il est évident que globalement c'est une liste qui recense les différentes manières dont quelque chose peut ne pas aller comme il faudrait [*go wrong*] avec ce qui apparaît vraisemblablement comme les structures de l'esprit façonnées par l'histoire biologique [*with the seemingly designed features of the mind*]. Très grossièrement, les troubles psychotiques impliquent des défaillances des processus de la pensée qui ne fonctionnent pas comme ils ont été façonnés à le faire par l'histoire biologique [*as designed*]; les troubles de l'anxiété impliquent des défaillances de l'anxiété et des mécanismes générateurs de la peur à fonctionner comme ils ont été façonnés à le faire par l'histoire biologique [*as designed*]; les troubles dépressifs impliquent des défaillances de la tristesse et des mécanismes de régulation aux réponses de perte [*loss-response*]; les comportements perturbateurs (troubles) de l'enfant impliquent des défaillances des processus de socialisation et des processus qui soutiennent la conscience et la coopération sociale; les troubles du sommeil impliquent une défaillance des processus du sommeil à fonctionner correctement; les dysfonctions sexuelles impliquent des défaillances de différents mécanismes relatifs à la réponse et la motivation sexuelles; les troubles de l'alimentation impliquent des défaillances des mécanismes de l'appétit, etc. Il y a une certaine quantité de non sens dans le DSM et ses critères sont souvent trop inclusifs. Néanmoins, la grande majorité des catégories sont suscitées par des états [pathologiques] que même une personne profane reconnaîtrait correctement comme des défaillances du fonctionnement façonné par l'histoire biologique [*failure of designed functioning*]²⁸. (WAKEFIELD 2007, p. 152.)

Cette citation découvre nous semble-t-il toute l'ambivalence de la position de Wakefield. D'un côté, celui-ci insiste sur la grande évidence de la réalité des troubles qu'il recense, n'hésitant pas à affirmer que « même une personne profane reconnaîtrait correctement » le bien-fondé des étiquettes du DSM-IV. Cette évidence, Wakefield l'a toujours assumée comme telle, considérant d'ailleurs que le terme de « dysfonctionnement » s'origine dans cette intuition vulgaire que « quelque chose ne

28. Nous avons décidé de traduire « *designed* » par « façonné par l'histoire biologique », étant donné la référence explicite et assumée par Wakefield à la théorie de l'évolution. Nous avons repris par ailleurs les traductions officielles en français des étiquettes du DSM-IV-TR auquel fait manifestement référence l'auteur tout au long de ce paragraphe.

va pas dans la personne» (*something has gone wrong "in the person"*)²⁹. Mais d'un autre côté, il insiste à plusieurs reprises dans son travail sur le fait que l'inférence dont il s'agit n'est qu'«hypothétique», qu'elle est «risquée³⁰» au sens où elle peut être falsifiée, et qu'elle doit être conduite prudemment en fonction de l'ensemble des preuves indirectes (*circumstantial evidence*)³¹ dont on dispose. Remarquons au passage que l'enjeu ici, pour la psychiatrie, est de défendre le bien-fondé de ses étiquettes actuelles après avoir consacré l'unité de sa tradition. Même si dans l'état actuel des choses, la classification des troubles mentaux ressemble davantage à un «guide de dépannage³²» qu'à une véritable nomenclature médicale, l'approche dite de la «boîte noire essentialiste» semble autoriser le type d'inférence que réalise spontanément le sens commun et les cliniciens, en visant un dysfonctionnement essentiel sous-jacent aux manifestations cliniques quand bien même l'essence de ce dysfonctionnement demeure indéterminée. Wakefield va jusqu'à vanter l'œcuménisme salutaire d'une telle approche, puisque, assure-t-il, cette indétermination de la boîte noire permet de transcender les oppositions théoriques qui minent la psychiatrie. Mais cet œcuménisme

29. WAKEFIELD (1992a, p. 240): "Basic to the concept of disorder put forward in DSM-III and DSM-III-R is the notion that a disorder is a condition in which something has gone wrong 'in the person'."

30. WAKEFIELD (1999a, p. 376): "I argue here that failure of a naturally selected function is necessary for disorder. This is a highly risky claim: it can be falsified by just one clear example of a disorder that is not an evolutionary dysfunction." Ou encore WAKEFIELD (1999b, p. 967): "The attribution of disorder ultimately involves a broad theoretical hypothesis that the cause of the symptoms involves a dysfunction, and this hypothesis can be falsified."

31. WAKEFIELD (2003, p. 971): "Obviously, in our present state of ignorance, we judge that there is a failure of some internal mechanism from circumstantial evidence that makes such a hypothesis overall the most plausible, just as, long before we understood etiology, we correctly recognized blindness and paralysis as disorders based on the circumstantial evidence of an individual's inability to see or move under conditions in which it was presumed that normally functioning individuals would be able to see or move, and where alternative explanations seemed unlikely."

32. WAKEFIELD (1999b, p. 971): "[...] The DSM's logical structure is closer to that of a trouble-hoarding guide than to that of a theory or research program."

de bon aloi et cette indétermination théorique sont-ils épistémologiquement satisfaisants ?

3.3 ➔ Les eaux troubles de l'inférence essentialiste...

À chaque fois que Wakefield mobilise son approche de la «boîte noire essentialiste», il fait appel à un même exemple pour illustrer son propos : l'exemple de l'eau. Cela se produit au moins à quatre occasions dans quatre articles différents : dans la discussion qu'on a évoquée précédemment autour du concept de fonction en vue de répondre à Millikan et Neander ; dans sa critique de l'essentialisme trop restrictif du psychologue Eysenck (WAKEFIELD 1997) ; dans une mise au point sur la compatibilité de sa position avec celle de Hacking (WAKEFIELD 2003, p. 981 et suiv.) ; enfin, dans une réflexion critique sur les thèses de Foucault (WAKEFIELD 2002). À chaque fois, l'enjeu est très différent : tantôt, c'est pour justifier l'unité essentielle du concept commun et du concept étiologique de fonction ; tantôt, c'est pour mettre en lumière le rapport causal qui s'exerce entre un dysfonctionnement sous-jacent et la symptomatologie de surface d'un trouble ; tantôt, c'est au contraire pour insister sur l'écart entre l'essence fixe du trouble et la malléabilité de ses manifestations cliniques ; tantôt enfin, c'est pour justifier la tendance naturelle de l'esprit humain à procéder à des inférences essentialistes.

Dans tous les cas, Wakefield s'appuie sur l'exemple paradigmatique de l'eau, c'est-à-dire d'une substance naturelle dont on sait identifier les propriétés de surface (*Le Petit Robert* les résume assez bien en présentant l'eau comme un «liquide incolore, inodore, transparent et insipide quand il est pur»), et dont la science n'a su caractériser que tardivement l'essence (gloire à Lavoisier et la formule chimique H_2O). On voit où porte l'analogie avec le trouble mental : si les propriétés de surface sont les symptômes cliniques du trouble mental, tout nous porte à inférer l'existence d'un dysfonctionnement qui explique en profondeur le trouble et donne sa cohérence au tableau clinique, quand bien même la science ne sait pas encore identifier définitivement ce dysfonctionnement. L'exemple de l'eau fournit le paradigme d'une telle inférence légitime :

Même les notions scientifiques les plus simples, comme par exemple l'idée qu'il y ait une substance «eau» qui peut être distinguée des autres substances par une certaine propriété interne structurale, et que l'eau liquide, la glace ou la vapeur sont toutes des instances de cette même

substance, [...], cela repose sur une inférence portant sur l'existence de structures causales sous-jacentes qui expliquent les [traits] apparents bien longtemps avant que la structure sous-jacente ne puisse être identifiée. (WAKEFIELD 1999b, p. 987.)

Cette analogie entre les concepts d'eau et de trouble mental, reprise maintes fois, est à prendre au sérieux. La seule nuance que Wakefield apporte, c'est lorsqu'il critique l'essentialisme comportementaliste d'Eysenck. Il affirme alors que le concept de trouble mental, contrairement à celui d'eau, n'est pas *purement* un concept essentialiste : c'est, comme il l'appelle, un concept «cause-effets», au sens où sa définition dépend *aussi* des propriétés de surface, puisque ne peut être trouble mental qu'un dysfonctionnement qui a des effets préjudiciables sur le plan clinique. De même qu'on n'est criminel que lorsqu'on a commis un crime, un dysfonctionnement, en vertu de la HDA, n'est un trouble mental qu'à condition d'être préjudiciable (*harmful*)³³. Mais pour le reste, l'analogie n'est pas seulement illustrative : le processus inférentiel dans les deux cas est le même. Il s'inscrit dans la même logique de découverte scientifique, il tolère les mêmes variations circonstancielle au niveau des traits de surface³⁴, et il est universellement partagé³⁵.

Pourtant, le peu de recul que Wakefield prend par rapport à cette analogie indique le point précis où sa théorie est faible. Car il faudrait bien commencer par là : cette analogie est-elle pour le moins fondée ? Entre l'application de la «boîte noire essentialiste» à un élément naturel comme l'eau et son application à un ensemble de conditions qui, dans

33. Alors qu'on aurait de bonnes raisons, selon Wakefield, de continuer à appeler «eau» une substance ayant la propriété chimique H_2O , mais ne présentant pas les propriétés de surface qu'on lui reconnaît habituellement (par exemple, à l'occasion d'une découverte sur une autre planète de l'eau à un état encore inconnu).

34. C'est l'enjeu de la discussion autour de Hacking, philosophe notoire pour sa volonté de concilier naturalisme et constructivisme. De même que l'eau et la glace, Wakefield écrit-il, sont une même substance dont les propriétés de surface varient en fonction des circonstances (comme la température), de même les manifestations cliniques d'un trouble peuvent varier suivant les circonstances (sociales ou culturelles) sans remettre en cause l'unité ou la réalité du trouble.

35. Tel est l'enjeu de l'argumentation dans "Fixing a Foucault sandwich" de 2002.

l'histoire humaine, ne s'est jamais imposé avec l'évidence que Wakefield souvent lui prête, il y a une différence qui devrait inviter à un minimum de circonspection. Or Wakefield ne met jamais véritablement en doute l'éventualité que chaque trouble mental identifié puisse être autre chose qu'une espèce naturelle (*natural kind*). Assurément, sa théorie *présuppose* qu'un trouble mental soit quelque chose comme un *natural kind*. Mais ni le fait que l'homme, naturellement, est prompt à procéder à des inférences essentialistes sur toutes sortes de choses, ni le fait plus spécialement qu'il a spontanément tendance à considérer un ensemble d'états mentaux négatifs comme pointant vers l'idée que «quelque chose ne va pas dans la personne» ne sont des garanties satisfaisantes pour considérer que les troubles mentaux sont bien ce qu'il croit qu'ils sont. Au contraire, aurait-on même raison de penser. Bachelard, dans *La Formation de l'esprit scientifique*, s'accordait lui aussi à considérer que l'homme est prompt à essentialiser. Mais il tirait de ce constat et de l'histoire des sciences une conclusion exactement inverse de celle de Wakefield, à savoir que cela constitue souvent un obstacle pour la pensée scientifique davantage qu'une garantie de son avancement. Wakefield, à cet égard, semble justement tomber dans les travers de «l'obstacle substantialiste» que dénonce Bachelard – si récurrent aussi bien dans l'histoire de la chimie que de la médecine –, lequel consiste à croire que «*la substance a un intérieur*; mieux, qu'elle *est un intérieur*», et à accorder une confiance spontanée «aux lumières de l'intuition qui nous installe au *cœur du réel*» (BACHELARD 1977, p. 101). De fait, non seulement l'opposition entre «propriétés de surface» et «essence sous-jacente» est une opposition qui structure la pensée de Wakefield sans jamais être clarifiée sur un plan épistémologique ou ontologique, mais celui-ci affirme encore que la tendance naturelle à l'essentialisme est une «tendance cognitive innée» qu'il faut d'une certaine manière assumer: «Quoi qu'il en soit, si nous avons tendance à être essentialistes par nature, écrit-il, alors le mieux c'est encore de se confronter à ce fait [*confront that fact*]» (WAKEFIELD 2002, p. 27). Le problème, Wakefield l'évoque comme une évidence sans trop s'y arrêter, c'est que «bien sûr, l'essentialisme peut nous égarer» (*ibid.*, p. 26).

Que l'essentialisme puisse nous égarer, c'est pourtant là le point fondamental. Car qu'est-ce qui nous garantit que les traits de surface convergent toujours vers une essence sous-jacente? Qu'est-ce qui légitime sur le plan scientifique le bien-fondé de nos inférences essentia-

listes? En focalisant sur cet exemple idéal de l'eau, ne risque-t-on pas de négliger d'autres exemples plus délicats, bien plus polémiques, mais plus proches aussi du type de danger que veut conjurer la psychiatrie, comme par exemple les inférences hasardeuses qui ont pu conduire certains à conclure d'un ensemble de propriétés de surface (taille et forme du crâne, physionomie, etc.) une nature commune au meurtrier ou une infériorité intellectuelle caractéristique de certaines races? Pourquoi un tel risque n'est-il pas sérieusement pris en compte? Tant que l'on n'a pas de confirmation par la science, l'inférence essentialiste reste une hypothèse, et sa comparaison avec d'autres inférences fructueuses du passé ne suffit assurément pas à en garantir le bien-fondé. Wakefield, sur ce point précis, est comme nous l'avons dit très ambivalent. Sa position ne cesse d'osciller entre une version faible de sa théorie, qui consiste à reconnaître l'utilité et même la nécessité pour la psychiatrie de se contenter d'inférences qui ne sont qu'hypothétiques, et une version forte où il s'agit de consacrer définitivement le bien-fondé des inférences à la base des étiquettes psychiatriques, dont il ne voit pas ce qui pourrait les remettre en cause. Dans un article de 2006, Wakefield semblait pourtant conscient de la difficulté qu'il devait affronter :

Il ressort clairement des comptes rendus historiques et anthropologiques que les valeurs, les normes et les idéologies influencent profondément ce que les gens croient être des fonctions naturelles, en particulier lorsque la compréhension scientifique de ce qui est fonctionnel ou dysfonctionnel fait défaut (comme c'est actuellement le cas pour de nombreux aspects mentaux). (WAKEFIELD 2006, p. 43.)

De deux choses l'une : ou bien Wakefield considère que la compréhension scientifique de ce qui est fonctionnel ou dysfonctionnel fait actuellement défaut en psychiatrie ; ou bien il considère que l'approche évolutionniste est aujourd'hui susceptible de combler ce manque. En d'autres termes, ou bien la HDA n'apporte rien d'autre que cette antique conviction fragile, ouverte à toutes les dérives, que « quelque chose ne va pas dans la personne ». Ou bien, elle prétend apporter une confirmation plus solide en s'appuyant sur les apports de la psychologie évolutionniste. Quelle est la position de Wakefield face à ce dilemme ? Du point de vue générique, la conception évolutionniste de la fonction est, nous l'avons dit, clairement présentée comme cette découverte scientifique qui aujourd'hui permet de légitimer rétrospectivement l'inférence intuitive fondatrice de la psychiatrie concernant la nature

du concept de trouble mental, comme la chimie jadis le fit pour l'eau. Une telle conclusion nous semble pour le moins hâtive, mais passons. Ce qui nous intéresse plutôt, ce sont les moyens théoriques employés à discriminer de manière effective les phénomènes normaux des phénomènes pathologiques. Or du point de vue des étiquettes spécifiques de la psychiatrie, Wakefield reconnaît que chaque inférence particulière n'est qu'hypothétique, appuyée au mieux sur l'ensemble des «preuves indirectes» (*circumstantial evidence*) dont nous disposons: «La HDA est une analyse du concept de trouble, pas une théorie des mécanismes ou des dysfonctionnements qui sous-tendent les troubles» (WAKEFIELD 2003, p. 978). Et pourtant, elle fournit «un cadre pour construire une théorie» (*ibid.*) particulière de chaque trouble mental. L'exemple privilégié et le plus approfondi que Wakefield ait présenté jusqu'à aujourd'hui est celui de la dépression, qu'il fait reposer, sur «la base de données transculturelles, développementales, comparatives et autres (par exemple, le deuil, l'attachement, l'expression faciale, etc.)» (cf. surtout HORWITZ & WAKEFIELD 2007), sur un dysfonctionnement de la tristesse entendue comme un mécanisme naturel de gestion des réponses aux phénomènes de perte. Dans ce cas précis de la dépression, Wakefield se sent tenu de détailler les observations et les raisons qui rendent son hypothèse plausible. Mais cela ne l'empêche pas, en d'autres endroits et pour d'autres troubles, de conclure avec une assurance qui ne peut que nous étonner :

Néanmoins, la plupart des catégories du DSM représentent des défaillances dans des fonctions pour lesquelles le *design* naturel typique de l'espèce n'est pas sérieusement mis en question [*the species-typical designed nature is not seriously in dispute*], comme pour le désir sexuel, le sommeil, la peur, la tristesse, la pensée, la motivation, etc.³⁶ (WAKEFIELD 1999b, p. 986.)

L'évidence qui consiste à justifier les étiquettes psychiatriques sur la base d'une référence à des facultés de l'esprit aussi vagues et englobantes que «la pensée» et «la motivation» est pour le moins embarrassante.

36. Cette citation est à mettre en contraste avec la suivante: "Whether dysfunction actually exists is an empirical issue not addressed by the HD analysis, which must be assessed case by case. If there is no reason to infer dysfunction, then the HD analysis disallows disorder attribution, contra DSM" (WAKEFIELD 2003, p. 971).

Le grave problème, c'est que Wakefield semble ne pas prendre la mesure du reproche qui lui a été adressé que le type d'inférence qu'autorise la HDA est, par son indétermination même, très ouvert à la possibilité de construire des théories *ad hoc* pour justifier le bien-fondé de n'importe quelle étiquette diagnostique. Wakefield lui-même n'est pas à l'abri d'un tel risque, et quelques exemples tirés de ses articles suffiront à montrer qu'il est bien réel. Le premier exemple s'inscrit dans la discussion d'une étiquette introduite dans le DSM-III-R qui a suscité une extraordinaire polémique, le trouble oppositionnel avec provocation (*Oppositional Defiant Disorder*). Wakefield fait la remarque suivante :

À la place de la définition opérationnelle et statistique clairement invalide du DSM-III-R du trouble oppositionnel avec provocation, qui étiquette les enfants comme atteints d'un trouble sur la base d'un comportement négatif supérieur-à-la-moyenne, on pourrait décider que l'attribution du trouble oppositionnel avec provocation est fondée sur une inférence qui part du comportement et qui porte sur un dysfonctionnement dans certains mécanismes inhibiteurs de l'agression (une identification vague est acceptable tant qu'on ne dispose pas de connaissances plus précises). (WAKEFIELD 1993, p. 170-171.)

En étant charitable, on pourrait dire que Wakefield ne cherche ici qu'à «sauver les phénomènes» (en l'occurrence la description clinique du trouble), en réorientant la description et la recherche vers la seule hypothèse qui à ses yeux soit légitime relativement à la HDA. Mais il est difficile de voir dans la redéfinition qu'il propose autre chose qu'une pirouette rhétorique. Pire, on peut se demander si en procédant ainsi, Wakefield n'immunise pas l'étiquette en question de toute critique possible en la rendant désormais difficilement falsifiable : quelle évidence empirique actuelle pourrait remettre en cause une telle hypothèse ? Et quel âge d'or de la science est-on sommés d'attendre qui puisse trancher clairement la question ? Prenons un deuxième exemple, tiré du même article de 1993, à savoir cette caractérisation étonnante de l'hypochondrie que Wakefield présente tout à fait incidemment dans le cours de son argumentation :

L'utilisation rationnelle de l'information est une fonction naturelle cardinale de l'esprit. Si les mécanismes rationnels ne fonctionnent pas correctement, alors un trouble peut être inféré même quand l'ignorance est un facteur causal. Par exemple, certains patients hypochondriaques ressentent une détresse irrationnelle à l'idée d'avoir certaines maladies jusqu'à ce qu'ils obtiennent des résultats de laboratoires qui

démontrent clairement qu'ils n'ont pas de maladie. Le manque d'une telle information probante est, alors, une condition causale nécessaire à leur détresse, mais cela n'implique pas que leur condition ne soit pas un trouble, car la relation entre l'information qui manque et l'anxiété qui en résulte est irrationnelle et implique que les systèmes cognitifs et affectifs ne fonctionnent pas correctement. (*Ibid.*, p. 168.)

Non seulement Wakefield redéfinit allégrement la théorie et la clinique de l'hypochondrie³⁷, mais il le fait sur la base d'une hypothèse qui pour le coup ouvre la voie à toutes les dérives possibles. En ce sens, toute anxiété forte et disproportionnée sera en tant que telle pathologique et révélera un dysfonctionnement physiologique sinon mental. Peut-être même faudra-t-il considérer comme pathologique tout raisonnement irrationnel mettant en péril cette «fonction naturelle cardinale de l'esprit» qui est de traiter rationnellement l'information. Dans tous les cas, justifier de la possibilité que l'hypochondrie soit une pathologie mentale sur l'hypothèse on ne peut plus vague d'un dysfonctionnement «des systèmes cognitifs et affectifs», cela n'est effectivement pas très différent que de la justifier plus simplement par la conviction que «quelque chose ne va pas dans la personne comme cela devrait aller».

Le dernier exemple, le plus illustratif sans doute, concerne le diagnostic de trouble de l'orgasme chez la femme. Prenant la défense du DSM-III-R, Wakefield considère dans un article de 1992 que les critères retenus par la classification, lorsqu'ils sont remplis, «arrivent presque à démontrer qu'il y a bien un dysfonctionnement derrière l'absence d'orgasmes³⁸». Le DSM-III-R, selon lui, parvient clairement à distinguer un tel trouble pathologique, qui implique un dysfonctionnement des «mécanismes internes de l'orgasme» de la femme (*her internal orgasmic*

37. Traditionnellement, l'hypochondrie est liée à la névrose, c'est-à-dire à un type d'affection qui n'a que peu à voir avec les mécanismes rationnels du traitement de l'information. Par ailleurs, cliniquement, on considère qu'il y a hypochondrie justement lorsque le patient continue d'être convaincu de l'existence de sa maladie même lorsque les examens demandés se sont révélés rassurants.

38. WAKEFIELD (1992a, p. 244) : "Despite lack of guidance from the definition of disorder, DSM-III-R criterion for inhibited female orgasm manages to rectify Masters and Johnson's (1970) error. It does so by adding a series of requirements that, if met, come close to demonstrating that there is indeed a dysfunction behind the lack of orgasms."

mechanisms) (*ibid.*), de la simple absence d'orgasmes qui peut être liée, comme chacun sait, à de nombreux facteurs extérieurs. Le caractère franchement fantaisiste de cette inférence essentialiste et surtout l'assurance avec laquelle elle est présentée laissent songeur quand on sait l'accent que Wakefield met sur le caractère «factuel» de sa définition. D'autant que l'ironie veut que les hypothèses évolutionnistes actuelles convergent vers l'idée que l'orgasme féminin a bien peu de chance d'avoir jamais été un effet sélectionné au cours de l'évolution (cf. LLOYD 2005), ce qui impliquerait qu'il n'a jamais été une fonction naturelle. Cet exemple soulève au passage une incongruité théorique de la HDA. Imaginons que Wakefield reconnaisse, sur l'évidence des travaux actuels, s'être trompé. Cela signifierait que ce qu'il considérait auparavant comme un trouble mental digne d'intérêt cesse, du jour au lendemain, d'avoir à être considéré comme tel, et doive logiquement être supprimé de la classification des troubles mentaux. Cela revient donc à dire que pour une patiente présentant exactement le même tableau clinique, avec exactement la même souffrance personnelle et psychique, ce dont dépend en dernière instance la légitimité de sa prise en charge psychiatrique, c'est la découverte que l'orgasme pour les femmes du pléistocène aurait pu constituer ou non un avantage sélectif. C'est absurde³⁹.

Il serait injuste de réduire l'ensemble des travaux de Wakefield, qui est si pertinent sur bien des sujets, à ces trois exemples litigieux. Nous pensons néanmoins que ces trois exemples permettent sur le plan théorique de pointer les insuffisances de son approche essentialiste. Quand elle ne fait pas directement appel à l'intuition, cette approche s'appuie sur des hypothèses évolutionnistes dont on peut craindre qu'elles soient ou bien difficilement falsifiables, ou bien grotesquement cousues pour l'occasion, ou bien tout simplement de peu de pertinence pour la psychiatrie. Quoi qu'il en soit, les connaissances factuelles manquent qui nous permettraient d'avancer dans ce domaine avec un minimum d'assurance. Aussi est-il à craindre que sur le plan méthodologique, la «boîte noire essentialiste» ne puisse apporter rien d'autre à la psychiatrie

39. Comme l'indique l'examen des «expérimentations conceptuelles» auxquelles se livre Wakefield dans un texte de 1999, ce qui dans sa théorie décide toujours, en dernier ressort, du caractère fonctionnel d'un trait, c'est bien l'histoire évolutive. Wakefield sur ce point est proche de la position de Millikan, sans vouloir en assumer les conséquences. Cf. WAKEFIELD (1999a).

qu'une «retraite dans l'obscurité⁴⁰». Loin de travailler à la clarification du statut scientifique des étiquettes psychiatriques, la HDA ne fait que renforcer une conviction médicale intuitive. L'approche évolutionniste dont Wakefield se réclame est trop permissive, et avec ou sans elle, sa théorie se résume à ce constat dressé par Dominic Murphy et Robert Woolfolk :

Si Wakefield considère ses étiquettes juste comme des pantonymes⁴¹ désignant n'importe quelle base interne qu'un trouble puisse avoir, alors sa théorie est bien plus modeste qu'elle n'y paraît de prime abord. Sa théorie devient plus plausible, mais beaucoup moins intéressante. Elle consisterait seulement à affirmer que s'il y a du pathologique, c'est que quelque chose d'indéterminé ne fonctionne plus comme il faut dans l'esprit. (MURPHY & WOOLFOLK 2000, p. 249.)

Tout nous porte à croire en effet que la HDA ne saurait aller plus loin dans la détermination du normal et du pathologique. Pourtant elle ne s'en tient pas là, puisqu'elle prétend apporter un éclairage, au moins partiel, sur l'*essence* de ce qui dysfonctionne. Le point précis de notre critique n'est que celui-ci : là où Wakefield affirme avec assurance que les explications fonctionnelles peuvent «être plausibles et très utiles, même lorsqu'on en sait très peu sur la nature réelle du mécanisme» (WAKEFIELD 1992b, p. 382), nous voulons au contraire rappeler que les explications fonctionnelles peuvent être trompeuses et très pernicieuses pour la science, surtout lorsqu'on en sait très peu sur la nature réelle des mécanismes et que l'on se repose sur des intuitions pour les établir. Il n'y a pas si longtemps, le génial Descartes, fort de toute son évidence, avait conclu que la fonction du cœur était de réchauffer le sang et de le raréfier. Et ce n'était pas faute d'aller y regarder de près, ni d'être accoutumé à «distinguer les vraies raisons des vraisemblables».

4 ⇨ Conclusion

Il ne suffit malheureusement pas de confier à la science le soin de trancher des questions litigieuses pour qu'elle soit en mesure de le faire. La fragilité des approches objectivistes de Boorse et de Wakefield, c'est

40. L'expression est de Houts. Cf. WAKEFIELD (2003, p. 982).

41. En littérature, un pantonyme (un tenant-lieu ou *placeholder* en anglais) est un terme très général et vague qui englobe d'autres termes sans les préciser (comme les mots «chose», «truc», «machin»).

d'être des approches «factuelles» qui laissent dans une grande indétermination épistémologique le type de «faits» sur lequel elles entendent s'appuyer. Nous avons cherché à montrer que cette indétermination ne tenait pas seulement au caractère incomplet de nos connaissances actuelles en psychiatrie, ni au nécessaire recul d'une théorie philosophique qui précisément laisse le soin à la science de fournir la matière au progrès, se contentant d'en indiquer la forme. Elle réside bien plutôt au cœur même de deux théories qui laissent indécisées les conditions de possibilité d'une démarcation objective du normal et du pathologique. Boorse a bien montré que la psychiatrie avait de bonnes raisons de prétendre au statut de branche autonome du savoir médical. Il n'est malheureusement pas parvenu à fournir un critère de démarcation intéressant qui soit susceptible d'éclairer les controverses psychiatriques. Wakefield a voulu être plus décisif sur ce dernier point, mais c'est au prix d'une inférence essentialiste qui, en faisant passer des assignations fonctionnelles problématiques pour des évidences, est loin d'avoir prouvé toute sa pertinence pour la psychopathologie.

Le discours fonctionnel en psychiatrie, qui s'est nourri principalement d'un débat importé de la philosophie de la biologie, n'est peut-être pas condamné à demeurer dans un tel état d'indétermination. Seulement, il faut bien considérer que son utilité actuelle pour la psychiatrie n'est que négative : il permet correctement de penser ce que la psychiatrie ne doit pas être. Mais il ne lui offre assurément pas les moyens, dans l'état actuel des connaissances, de conforter le caractère scientifique de son entreprise. Exiger davantage de lui, c'est courir le risque de remplacer l'excès des antipsychiatres, qui soupçonnaient systématiquement l'exercice d'une police insidieuse des conduites derrière toute étiquette psychiatrique, par l'excès contraire, celui de surestimer le poids scientifique des intuitions naturalistes qui se trouvent derrière la grande majorité des étiquettes établies par les psychiatres. Face à ce problème, peut-être le mieux est-il de renoncer tout bonnement à la perspective de fixer un critère biologique général et définitif du normal et du pathologique valable pour l'ensemble des troubles mentaux. Peut-être la réflexion philosophique sur la psychiatrie gagnera-t-elle à se concentrer davantage sur la complexité des raisons toujours singulières et régionales, tant cliniques, scientifiques que sociopolitiques, qui motivent l'inclusion d'une souffrance psychologique donnée dans le long cortège des troubles mentaux.

☉Remerciements☉

Je remercie particulièrement, pour leurs remarques et leurs commentaires, Jean Gayon, Élodie Giroux, Françoise Longy, Luc Faucher et Philippe Huneman.

☉Références bibliographiques☉

A

ANANTH M. (2008), *In Defense of an Evolutionary Concept of Health: Nature, Norms, and Human Biology*, Ashgate Studies in Applied Ethics.

B

BACHELARD G. (1977), *La Formation de l'esprit scientifique*, Vrin.

BERNARD C. (1984), *Introduction à l'étude de la médecine expérimentale* [1865], Champs Flammarion.

BOORSE C. (1975), "On the Distinction Between Disease and Illness", *Philosophy and Public Affairs*, 5: 49-68.

BOORSE C. (1976), "What a Theory of Mental Health Should Be", *Journal of the Theory of Social Behavior*, 6 (1): 61-84.

BOORSE C. (1977), "Health as a Theoretical Concept", *Philosophy of Science*, 44(4): 542-573.

BOORSE C. (1997), "A Rebuttal on Health", in J. Humber & R. Almeder (eds.), *What is Disease?*, Humana Press.

C

CANGUILHEM G. (1998), *Le Normal et le pathologique*, PUF.

CUMMINS R. (1975), "Functional Analysis", *Journal of Philosophy*, 72: 741-760.

D

DAVIDSON D. (1963), "Actions, reasons, and causes", *The Journal of Philosophy*, 60: 685-700.

F

FULFORD K.W.M. (2001), "What is (mental) disease? An open letter to Christopher Boorse", *Journal of Medical Ethics*, 27: 80-85.

G

GIROUX É. (2009), «Définir objectivement la santé: une évaluation du concept bio-statistique de Boorse à partir de l'épidémiologie moderne», *Revue philosophique de la France et de l'étranger*, 134 (1): 35-58.

H

HORWITZ A.V. & WAKEFIELD J.C. (2007), *The Loss of Sadness. How Psychiatry Transformed Normal Sorrow Into Depressive Disorder*, Oxford University Press. Trad. fr. par F. Parot, *Tristesse ou dépression? Comment la psychiatrie a médicalisé nos tristesses*, Mardaga, 2010.

K

KENDELL R.E. (1975), "The concept of disease and its implications for psychiatry", *British Journal of Psychiatry*, 127: 305-315.

L

LLOYD E. (2005), *The Case of the female orgasm: bias in the science of evolution*, Harvard University Press.

M

MILLIKAN R. (1989), "In Defense of Proper Functions", *Philosophy of Science*, 56(2): 288-302.

MURPHY D. (2005), "Can evolution explain insanity?", *Biology and Philosophy*, 20: 745-766.

MURPHY D.P. & WOOLFOLK R.L. (2000), "The Harmful Dysfunction Analysis of Mental Disorder", *Philosophy, Psychiatry, and Psychology*, 7(4): 241-252.

N

NEANDER K. (1991), "Functions as selected effects: The conceptual analysis defense", *Philosophy of Science*, 58: 168-184.

S

SCHAFFNER K.F. (1993), *Discovery and explanation in biology and medicine*, University of Chicago Press.

SZASZ T. (1960), "The Myth of Mental Illness", *American Psychologist*, 15: 113-118.

W

WAKEFIELD J.C. (1992a), "Disorder as harmful dysfunction: A conceptual critique of DSM-III-R's definition of mental disorder", *Psychological Review*, 99: 232-247.

WAKEFIELD J.C. (1992b), "The concept of mental disorder: On the boundary between biological facts and social values", *American Psychologist*, 47: 373-388.

WAKEFIELD J.C. (1993), "Limits of operationalization: A critique of Spitzer and Endicott's (1978) proposed operational criteria for mental disorder", *Journal of Abnormal Psychology*, 102: 160-172.

WAKEFIELD J.C. (1997), "Diagnosing DSM-IV, Part 2: Eysenck (1986) and the essentialist fallacy", *Behavior Research and Therapy*, 35(7): 651-665.

WAKEFIELD J.C. (1999a), "Evolutionary versus prototype analyses of the concept of disorder", *Journal of Abnormal Psychology*, 108: 374-399.

WAKEFIELD J.C. (1999b), "Philosophy of science and the progressiveness of the DSM's theory-neutral nosology: response to Follette and Houts, Part 1", *Behaviour Research and Therapy*, 37: 963-999.

WAKEFIELD J.C. (2000), "Aristotle as sociobiologist: The 'function of a human being' argument, black box essentialism, and the concept of mental disorder", *Philosophy, Psychiatry, and Psychology*, 7(1): 17-44.

WAKEFIELD J.C. (2002), "Fixing a Foucault sandwich: Cognitive universals and cultural particulars in the concept of mental disorder", in K.A. Cerulo (ed.), *Culture in mind: Toward a sociology of culture and cognition*, Routledge: 245-266.

WAKEFIELD J.C. (2003), "Dysfunction as a factual component of disorder: Reply to Houts, Part 2", *Behavior Research and Therapy*, 41: 969-990.

WAKEFIELD J.C. (2006), «Fait et valeur dans le concept de trouble mental: le trouble en tant que dysfonction préjudiciable», *Philosophiques*, 33(1): 37-64.

WAKEFIELD J.C. (2007), "The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis", *World Psychiatry*, 6: 149-156.

WRIGHT L. (1973), "Function", *Philosophical Review*, 82: 139-168.

☞ Notice biographique ☞

Agrégé de philosophie, **Steeves Demazeux** termine une thèse de philosophie des sciences sur les classifications psychiatriques américaines sous la direction de Jean Gayon (Université Paris1). Il est doctorant à l'IHPST (Institut d'histoire et de philosophie des sciences et des techniques), attaché temporaire d'enseignement et de recherche (Ater) à l'Université Paris 1, et participe au programme ANR PHS2M

(Philosophie, histoire et sociologie de la médecine mentale) dirigé par Pierre-Henri Castel (Cesames-CeRMes 3).



Avertissement. Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur et des auteurs. Les textes et illustrations figurant sur ce site peuvent être consultés et reproduits sur un support papier ou numérique sous réserve qu'ils soient strictement réservés à un usage personnel, scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document selon ce modèle :

Steeves Demazeux, «Le concept de fonction dans le discours psychiatrique contemporain», in Gérard Lambert et Marc Silberstein (dir.), Matière première. Revue d'épistémologie [en ligne]. Nouvelle série, N° 1/2010 : Épistémologie de la médecine et de la santé, Éditions Matériologiques. Mis en ligne le 7 octobre 2010. URL : www.materiologiques.com

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.

Cyrille Bouvet **Causalité(s) en psychopathologie : des modèles inconciliables ?**

Il convient dans un premier temps d'expliquer ce titre et d'en situer chacun de ses termes. La psychopathologie est une sous-discipline de la psychologie ayant pour objectif l'étude des troubles psychiques ainsi que des interventions psychothérapeutiques. Elle est dite «clinique» en ce que son unité d'étude est, en principe, le sujet, dans sa globalité et sa complexité psychique. Au sein des universités françaises, il existe souvent une différenciation (équipe, laboratoire) entre la psychologie clinique et la psychopathologie, la première se définissant par son approche du sujet (approche dite «clinique» faite d'écoute, d'empathie et de compréhension subjective), la seconde par son objet (les troubles psychiques). Mais cette différenciation est en partie artificielle : la psychopathologie s'affirme «clinique» dans ses méthodes et la clinique s'applique essentiellement à propos des troubles psychiques, comme son étymologie le suggère («au lit du malade»), même si, ici ou là, ses principes compréhensifs subjectifs s'appliquent aussi à des sujets sans troubles psychiques. Le manque de clarté des rapports entre psychologie clinique et psychopathologie n'étant pas levé par ces précisions, je considérerais ici que «clinique» et «patho» participent de la même sous-discipline de la psychologie, celle qui se donne pour objet la compréhension des troubles psychiques ainsi que des modèles d'intervention psychothérapeutiques. C'est ce à quoi fait référence le terme «psychopathologie» dans le titre.

Pour aller plus avant, il est aussi nécessaire de définir les troubles psychiques. À l'heure actuelle, le problème qu'ils posent est d'abord celui de leur identification et de leur description. S'il existe un accord global entre praticiens et chercheurs sur de grandes catégories de troubles, de nombreuses questions fondamentales restent posées. La première porte sur les limites entre le normal et le pathologique. Les troubles psychiques sont-ils des entités singulières comme l'est par exemple le virus de la grippe pour l'organisme qu'il attaque, auquel cas on pourrait dire sans hésitations qu'un sujet est ou n'est pas malade psychique et ainsi la frontière santé/maladie serait nette ; ou faut-il considérer que les troubles psychiques sont des phénomènes normaux ayant dépassé un certain seuil (un peu d'anxiété est normale, beaucoup est pathologique – mais comment fixer le seuil...) ? En d'autres termes, les troubles psychiques (dépression, schizophrénie, etc.) sont-ils des catégories (homogènes à l'intérieur, hétérogènes avec les autres catégories dont la santé) ou des dimensions psychiques (comme l'humeur, la gestion émotionnelle, l'ouverture relationnelle, etc.) s'exprimant de façon très variable selon les sujets ? Des arguments permettent d'abonder dans ces deux directions, catégorielle et dimensionnelle, sans pouvoir trancher nettement en faveur de l'une ou de l'autre. Un autre problème réside dans le repérage même des troubles, leur diagnostic. En l'absence de marqueur matérialisable (biologique, par exemple), les professionnels peinent parfois à situer un même patient dans une même catégorie (30 % de différence pour des schizophrènes, par exemple, dans des recherches empiriques et contrôlées – CONFÉRENCE DE CONSENSUS 1994). Toutes ces raisons ont conduit l'OMS à utiliser dans sa classification des maladies (CIM10) pour la partie psychiatrique le terme de «trouble» plutôt que celui de «maladie». Par ailleurs, ces questions ne se posent pas de façon identique suivant les troubles concernés tant ceux-ci sont divers et nombreux, allant d'une anxiété légère quasi normale aux troubles schizophréniques les plus désorganisateur. Il peut s'agir de peur irrationnelle (phobie), d'anxiété permanente et intense, de culpabilité envahissante, de symptômes dépressifs (fatigue, idées noires, aboulie), d'envahissement par des rituels obsessionnels (passer 8 heures par jour à se laver les mains), de difficultés relationnelles, de troubles du sommeil, des comportements alimentaires, des relations avec la société et la loi, d'hallucinations, de délire, de suicide, d'addiction à des produits (alcool, drogues, etc.), de l'interaction entre plusieurs de ces éléments

ou encore de l'absence de cela et de toute vie psychique dynamique liée à un sentiment de vide existentiel. Pour compléter cette liste non exhaustive, l'on peut aussi parler des troubles envahissants du développement (dont l'autisme) et des détériorations neurocognitives liées à l'âge (Alzheimer), qui ne sont pas sans conséquences graves sur l'équilibre psychologique.

Dans ce contexte, où l'effort d'identification est encore dominant, la question de l'origine des troubles, de leur causalité, peut sembler quelque peu prématurée. En effet, comment rechercher l'origine d'un phénomène mal défini? Pourtant, la construction d'hypothèses causalistes est nécessaire à l'élaboration de stratégies thérapeutiques, elles-mêmes indispensables à la prise en charge des personnes souffrant de troubles psychiques. Il y a donc un impératif de soin déterminant la construction de modèles de la causalité des troubles psychiques, malgré la fragilité du savoir à leur propos. La pratique soignante est en avant de la science.

1 ↻ La causalité des troubles psychiques

Pris dans son sens le plus immédiat, la causalité renvoie à un rapport entre deux phénomènes, l'un étant cause de l'autre, le premier étant suffisant et nécessaire pour qu'advienne le second, qui sera alors nommé «conséquence» ou «effet». Mais cette causalité simple et linéaire apparaît bien démunie pour expliquer la vie psychique. Pour saisir cette dernière il convient de complexifier les concepts et de parler de causalité multiple, complexe, circulaire, sans pour autant que ces derniers concepts rendent compte actuellement de façon absolument convaincante des causes des phénomènes psychiques en général et des troubles psychiques en particulier. Les troubles psychiques, par leur étrangeté, leurs conséquences handicapantes et les souffrances qu'ils infligent, ont de tout temps été sources de questionnements et de tentatives d'explications. L'explication venant combler le besoin de sens et de compréhension, et apportant à la société un guide d'action face aux troubles psychiques. Les modèles de causalité des troubles psychiques ont été nombreux dans l'histoire et le sont encore aujourd'hui. Les modèles occidentaux anciens furent des modèles physiologiques métaphoriques (l'hystérie causée par les mouvements de l'utérus), des modèles religieux (les hallucinations comme possession par une force surnaturelle, positive – Jeanne d'Arc – ou négative – les sorcières) ou encore des modèles moralistes (les troubles provenant d'un dérègle-

ment des mœurs). À chaque fois, ces modèles de causalité proposent une direction d'action (physiologique – des sels – quand la cause est supposée organique; de vénération ou d'élimination de l'individu quand il est supposé possédé; d'enfermement, de punition et de redressement moral quand il est considéré vicieux). Les modèles actuels sont développés par différentes disciplines des sciences médicales et humaines. Les neurologues s'attellent à comprendre l'influence des dérèglements dans les neurotransmissions (application: pharmacopée psychotrope); les généticiens recherchent les causes génétiques supposées (application: médecine génétique à venir?); les sociologues insistent sur la sociogénie des troubles (application: agir en faveur d'une société moins pathogène); etc. Parmi les modèles de causalité existant, l'on retrouve plusieurs modèles psychologiques, comme le modèle psychanalytique, ceux des théories de l'apprentissage et des théories cognitives, le modèle systémique. Ce sont ces trois modèles qui seront discutés plus bas.

Les modèles de causalité actuels des troubles psychiques vont du plus organique (modèle génétique, modèles infectieux – par exemple pour la schizophrénie; CONFÉRENCE DE CONSENSUS 2003) au plus social (modèles sociogénétiques – REICH 1999), en passant par des modèles psychologiques. Chaque modèle a des arguments convaincants pour défendre son point de vue, mais il ne peut le défendre que partiellement seulement. Actuellement, les troubles psychiques sont considérés comme relevant d'une causalité multiple associant des dispositions physiologiques héritées et/ou acquises (par exemple liées à la prise de produits actifs psychiquement), des éléments liés à l'éducation précoce (par exemple, être élevé par des parents dépressifs) et des éléments liés aux expériences tout au long de la vie (des traumatismes, des événements positifs, etc.), des facteurs liés à l'environnement actuel du sujet (la situation familiale, sociale, professionnelle), et, enfin, des facteurs liés à la façon dont le sujet traite psychiquement ces éléments et les traces qu'ils ont laissées en lui. Cette causalité s'inscrit donc dans une perspective «biopsychosociale» (ENGELS 1977)

La causalité des troubles est donc considérée comme plurielle, mais les différents facteurs de causalité ne jouent probablement pas de la même façon ni avec la même intensité suivant le type de troubles. Pour certains de ces troubles l'hérédité semble jouer un rôle important (troubles bipolaires, autisme – LAZARTIGUES 2005) mais non exclusif; pour d'autres, les événements de vie semblent déterminants (syndrome de

stress post-traumatique, par exemple), mais non exclusifs non plus ; pour d'autres encore, la situation sociale actuelle semble déterminante (dépression réactionnelle à une exclusion professionnelle, par exemple), mais là encore pas exclusive. Par ailleurs, si l'on suppose ces différents facteurs à l'œuvre, les arguments allant dans ce sens sont essentiellement corrélationnels (mais cela n'indique pas un rapport de causalité seulement une coïncidence), cliniques (les récits des patients et des familles), ou théorique (la théorie suppose que...). Mais, à ce jour, il n'y a pas de preuves indiscutables confirmant tel ou tel modèle de causalité de tel ou tel trouble psychique, seulement des indices plus ou moins concordants. Les facteurs en jeu sont nombreux et leurs relations sont complexes et pour aucun trouble psychique nous n'avons actuellement de certitude quand à l'articulation de ces différents facteurs ni même parfois s'ils sont réellement en jeu et, quand ils le sont, s'ils sont une cause ou une conséquence des troubles. Cependant, le modèle général de causalité des troubles psychiques le plus souvent admis, au-delà de la diversité des disciplines, est le modèle stress/vulnérabilité qui a l'avantage d'intégrer des facteurs que l'on pourrait appeler extérieurs (les stressseurs de l'environnement : éducation, situation sociale, traumatisme, infection, etc.) et des facteurs individuels ou intérieurs (physiologiques, psychologiques, etc.) et cela dans un rapport interactif. Cette hypothèse large apparaît la plus cohérente avec les données disponibles.

Bref, pour aucun trouble psychique nous ne savons de façon certaine ce qui le cause. Et la conséquence de ce constat d'ignorance globale, c'est que pour aucun trouble psychique nous ne connaissons de façon certaine les facteurs concourant à sa diminution et à l'amélioration du patient. Ce qui n'empêche pas bon nombre d'approches thérapeutiques (médicales, sociales et psychologiques) d'avoir fait preuve de leur efficacité curative dans la plupart des troubles psychiques. Mais si l'on constate des améliorations des patients, parfois très importantes et définitives, les causes de celles-ci sont encore âprement discutées entre spécialistes et sont loin de faire consensus, tout comme les causes des troubles psychiques. Nous reviendrons sur ces questions dans la dernière partie de ce texte.

L'objectif ici est de questionner les modèles psychopathologiques des troubles psychiques, ainsi que leurs modèles d'interventions sous l'angle de leurs conceptions de la causalité (l'étiologie) des troubles. Comme on

l'a vu plus haut, la portée scientifique de ces modèles est assez incertaine à ce jour. Par ailleurs, il existe de nombreux autres modèles non psychologiques ainsi que pléthore de modèles et sous-modèles étiologiques psychologiques et de modèles d'intervention psychothérapeutique (on recense environ 400 psychothérapies différentes, dont plusieurs se réfèrent à des modèles étiologiques similaires). Il sera ici question des trois modèles principaux (les plus utilisés à ce jour) : psychanalytique, cognitifs et comportementaux, systémiques. Il est probable qu'aucun de ces modèles ne soit suffisamment puissant pour expliquer globalement tous les troubles psychiques. Pourtant, et cela depuis un siècle, en psychologie, ils ont tous eu, à un moment donné de leur histoire, une tendance théorique et thérapeutique optimiste les poussant à adopter une attitude intellectuelle et clinique dogmatique, dans le sens où chacun a pu prétendre abusivement expliquer l'ensemble du problème de l'étiologie et du soin (VIGOSTKI 1927). Il ne sera pas question ici d'entrer dans le détail de toutes les adaptations de chacun de ces modèles, mais de se centrer sur ce qui en fait le cœur, sur ce qui en est à l'origine et qui est actif dans chacun de leurs prolongements, toujours en suivant l'axe du problème de la causalité des troubles et des changements en psychothérapie.

Je conclurai en discutant du problème posé dans le titre : ces modèles sont-ils conciliables et si oui, à quelles conditions peut-on envisager un modèle intégratif ?

2 ⇨ Psychanalyse : une quête vertigineuse de la causalité première

Comme le soulignent bon nombre de psychanalystes (BOTELLA 2001), il est maintenant difficile de parler de la psychanalyse tant les conceptions théoriques et techniques s'en revendiquant sont diverses. Parler *des* psychanalyses serait sans doute plus approprié. C'est pourquoi il ne sera question ici que de la psychanalyse freudienne classique telle que Freud la développe dans la première partie de son œuvre (approximativement jusqu'en 1914), et qui inclut la première topique et la première théorie des pulsions. Complétée et remaniée ultérieurement par Freud, cette théorisation reste malgré tout aujourd'hui encore déterminante. On va voir que la question de la causalité y tient une place très importante. En effet, la psychanalyse naît par la recherche par Freud des

causes des névroses et en particulier de la névrose hystérique. Cela apparaît nettement dans la citation suivante :

Une observation fortuite nous a amenés à rechercher depuis quelques années déjà, dans les formes et les symptômes les plus divers de l'hystérie, la cause, l'incident, qui a, pour la première fois et souvent très loin dans le passé, provoqué le phénomène en question. (FREUD 1892, p. 1.)

Une autre citation, dans le même ouvrage, nous éclaire sur les raisons de cette quête de la cause :

À notre très grande surprise, nous découvrîmes, en effet, que chacun des symptômes hystériques disparaissait immédiatement et sans retour quand on réussissait à mettre en pleine lumière le souvenir de l'incident déclenchant, à éveiller l'affect lié à ce dernier et quand ensuite le malade décrivait ce qui lui était arrivé de façon fort détaillée et en donnant à son émotion une expression verbale. (*Ibid.*, p. 4.)

Dans ces citations de la période prépsychanalytique, on voit donc le projet de base de la psychanalyse se dessiner clairement et déterminer ses développements ultérieurs. Ce projet est donc très concerné par le problème de la causalité et cela, finalement, à tous les niveaux. En effet, la recherche des causes détermine en même temps le projet théorique (les événements de la vie psychique sont déterminés par certains événements passés), l'objectif thérapeutique (la découverte des causes du trouble – et son expression verbale – est source de guérison), le projet technique : la méthode (cathartique d'abord, cure type ensuite) a pour objectif de faciliter la recherche des causes. De ce point de vue, on peut dire que le projet psychanalytique se ramasse pour l'essentiel autour de la question de la causalité.

Cela est très net dans les citations freudiennes ci-dessus, qui appartiennent à la théorisation prépsychanalytique et dont le pendant technique est la recherche de la catharsis (sous hypnose d'abord, par suggestion non hypnotique ensuite, par l'association libre enfin) consistant dans l'expression des souvenirs refoulés et de leurs affects afférents. La théorie freudienne prépsychanalytique est la théorie dite « de la séduction », ou « neurotica » qui explique les troubles hystériques. Son principe causaliste est le suivant : le sujet subit dans son enfance une agression sexuelle (viol, attouchements, exhibition, etc.) qu'il ne parvient pas à assimiler psychiquement. Il en refoule donc les traces psychiques hors de sa conscience, dans l'inconscient. Les troubles sont la manifestation

détournée des traces psychiques inconscientes de cet événement traumatique et sexuel. La thérapie a pour but, avec la « psycho-analyse », de rendre ces traces conscientes ce qui a un effet cathartique (libérateur), donc thérapeutique, selon ce modèle. La quête de la cause première (un événement passé réel) est donc au cœur de cette conception causale linéaire. En 1897, Freud affirme dans un courrier : « Je ne crois plus à ma neurotica » (FREUD 1887-1902), c'est-à-dire qu'il ne croit plus à la réalité de l'événement traumatique pour diverses raisons (sociales, cliniques, etc.). La théorie évolue donc pour l'essentiel vers l'internalisation de la cause première. Ainsi, le névrosé adulte n'aurait pas été agressé sexuellement par une personne extérieure à lui, mais il aurait eu, étant enfant, des fantasmes (désirs, craintes) sexuels inassimilables par sa conscience et les processus secondaires (fonctions psychologiques supérieures), et ces fantasmes sexuels auraient donc été refoulés. Cette internalisation de la cause sexuelle traumatique première signe la naissance de la psychanalyse car elle ouvre la voie à l'exploration de la « réalité psychique » inconsciente.

De ce fait, la technique psychanalytique (cure type, « divan-fauteuil », association libre, etc.) s'organise autour de cet objectif : l'exploration de l'inconscient. La cause des troubles hystériques (et finalement de la plupart des troubles psychiques, par extension) se trouve dans l'inconscient, dans les fantasmes sexuels refoulés. Ceux-ci cherchant à s'exprimer pour se satisfaire, les mécanismes de défense (refoulement, par exemple) s'efforcent de les contenir dans l'inconscient car ces désirs sexuels heurtent la morale sociale et familiale intériorisée par le sujet. Ce conflit peut se résoudre par des symptômes psychiques, qui sont l'expression du fantasme déformé et rendu méconnaissable par les mécanismes de défense. Finalement, le symptôme est un compromis entre le désir et la défense. On peut donc voir la causalité en psychanalyse sous cet aspect linéaire et cela selon plusieurs lignes directrices allant de l'inconscient vers le conscient, du profond vers le superficiel, du passé vers le présent... Ces lignes représentent le schéma théorique de base. La dynamique thérapeutique consiste à remonter la chaîne des causalités vers sa source, à découvrir dans le symptôme ce qui fut refoulé et dans ce qui fut refoulé ce qui fut désiré, en allant cette fois-ci du conscient vers l'inconscient, du présent vers le passé, du superficiel vers le profond.

Ce modèle causaliste linéaire correspond bien aux premières théories psychanalytiques et il est encore très répandu, en particulier chez les personnes qui ont une vision simplifiée de la psychanalyse. Cependant, ce serait réducteur d'y limiter l'ensemble de la théorisation psychanalytique freudienne. De fait, l'on peut aussi concevoir ses modèles de façon systémique et complexe quand on considère par exemple les topiques freudiennes (par exemple : le système moi/ça/surmoi dont les éléments correspondent à des fonctions et des «lieux» psychiques en interaction complexe – cf. LAGACHE 1961). Pour être complet, il faudrait considérer une causalité linéaire correspondant à la chronologie du trouble et une causalité circulaire (systémique) rendant mieux compte de la dynamique présente de l'appareil psychique. La cure psychanalytique freudienne a donc un double objectif : remonter de façon linéaire la chaîne des souvenirs et analyser le fonctionnement psychique actuel du sujet (l'organisation défensive en particulier) avec pour objectif de rendre plus efficace et adapté le fonctionnement de ce système psychique.

Quoiqu'il en soit, si l'on se penche sur la dimension linéaire de la causalité en psychanalyse, sur la fonction centrale de la quête de causalité comme fondement théorique, thérapeutique et technique, il est intéressant de constater les problèmes qu'elle posait et pose encore probablement.

Pour que cette théorisation tienne, il est nécessaire qu'il existe une première cause, un premier événement (externe ou interne) provoquant la succession de transformations des traces psychiques de cet événement (refoulement puis retour du refoulé), aboutissant finalement au symptôme. Il est intéressant d'étudier les différentes destinées théoriques du statut de cette cause première. Avant 1897, comme il est décrit plus haut, cet événement était considéré par Freud comme ayant eu lieu réellement. Il s'agissait donc d'un événement historique, qui pouvait être reconstruit par les souvenirs du patient, voire par les souvenirs de ses proches. Il était matérialisable et observable. Cependant, Freud note assez rapidement que derrière l'événement premier évoqué par le patient, ses souvenirs tendent à en proposer un autre encore plus ancien. Et régulièrement, les événements premiers sont détrônés par des souvenirs antérieurs sans qu'il semble y avoir de fin à ce mouvement de régression, à cette plongée vers le passé, à ce recul dans l'inconscient. C'est une des raisons qui a contraint Freud à abandonner sa «neurotica» et à situer la cause du trouble à l'intérieur de la psyché du sujet

(après, il est vrai, l'avoir situé dans les activités onanistes du sujet et la culpabilité conséquente).

À partir de là, plusieurs théorisations psychanalytiques vont tenter de situer l'origine psychique des fantasmes sexuels refoulés. La plupart ont comme caractéristique de dématérialiser l'événement premier et même de le sortir de l'histoire observable du sujet. La source première est d'abord conceptualisée par Freud avec le « complexe d'Œdipe ». Les symptômes psychiques proviendraient des aléas de ces désirs sexuels et haineux universels vis-à-vis des parents. Freud situe ce complexe dans la petite enfance vers 3 ans. Mélanie Klein, par la suite, situera ce complexe dès les premiers mois de vie extra-utérine, reculant d'autant plus la source première. D'autres auteurs, dans une perspective trans-générationnelle, situeront l'origine de ces fantasmes chez les éducateurs (parents) de l'enfant : leurs propres problématiques inconscientes laisseraient une empreinte profonde et inconsciente chez l'enfant qui serait à l'origine des fantasmes, eux-mêmes, par leur refoulement, causes des troubles. Freud va encore plus loin dans la dématérialisation, en affirmant que les fantasmes sexuels inconscients trouvent leur origine dans les fantasmes dits précisément « originaires » (FREUD 1915). Ces fantasmes, inconscients et universels, transmis héréditairement depuis l'origine de l'humanité seraient en fait la représentation inconsciente de situations vécues par les premiers hommes dont Freud retrace le mythe dans *Totem et tabou* (1912).

Ainsi, cette théorisation linéaire de la causalité voit la cause première située vers des zones de plus en plus inexploables (et donc invérifiables) qui sont l'histoire très précoce du patient (dont il n'a pas de souvenirs personnels), l'inconscient (le ça) du sujet, l'inconscient de ses parents et finalement, suivant une chaîne causale quasi infinie, la préhistoire psychique de l'humanité. Ainsi, de ce point de vue, la quête de la cause ultime est marquée par un toujours plus (plus tôt, plus profond, plus archaïque, plus inconscient) qui a conduit la psychanalyse sur des terrains situés hors du champ de la psychologie. Ce dernier constat est d'ailleurs revendiqué par Freud qui qualifie sa théorie de « métapsychologique ».

Cette fuite en arrière de la causalité première reléguée sur des terrains immatériels, inobservables et abstraits, éloigne les théories métapsychologiques de la démarche scientifique (dont les principes sont explicités dans LECOINTRE 2007). Cependant, dans sa richesse, le champ

psychanalytique contient aussi des démarches qui se maintiennent dans le champ psychologique et s'efforcent de répondre aux exigences scientifiques (par exemple, WESTEN 1999).

En suivant cette causalité linéaire psychanalytique, on en mesure l'importance théorique, thérapeutique et technique. Mais il apparaît aussi, en quelque sorte que trop de causalité tue la causalité. C'est-à-dire que la fuite en arrière de la quête de la cause première des fantasmes sexuels conduit parfois la théorie dans des spéculations invérifiables et par conséquent sans pouvoir explicatif et avec peu d'incidences pratiques et thérapeutiques (comment analyser des fantasmes originaires ?). Quoiqu'il en soit, pour agir sur la situation présente, la psychanalyse étend son regard causaliste très loin dans le passé.

3 ⇨ Théories de l'apprentissage et cognitivisme : les modèles des thérapies comportementales et cognitives (TCC)

Passons maintenant aux modèles des approches TCC qui sont en fait structurées par deux types de modèles différents : ceux des théories de l'apprentissage et ceux du cognitivisme. Chacun d'entre eux développent en son sein plusieurs variantes, mais nous ne verrons que les modèles de base.

3.1 ⇨ Le comportementalisme : repérer la cause dernière (ou déclenchante)

Les modèles du comportementalisme proviennent des théories de l'apprentissage développées au début du xx^e siècle. Les principales sont celles du conditionnement classique (ou répondant) de Pavlov, celle du conditionnement opérant de Skinner et enfin, celle de l'apprentissage social (ou vicariant) élaborées par Bandura.

Watson, entre les deux guerres, a été le héraut, certains disent le père, du comportementalisme en tant qu'il a envisagé l'application des données des théories de l'apprentissage et du conditionnement classique sur les conduites humaines en général. Naville a introduit la démarche de Watson en France (NAVILLE 1942).

Le comportementalisme s'attache à expliquer, puis prévoir, provoquer ou changer, les comportements des êtres vivants et parmi ceux-ci, des êtres humains. Les troubles psychiques sont considérés comme des comportements inadaptés et l'action thérapeutique a pour objectif de les modifier ou de les supprimer. Les troubles psychiques, comme tout

comportement, peuvent être ouverts (visibles, comme de l'évitement, des rituels, etc.) ou couverts (internes, comme de l'angoisse, des pensées, etc.). Contrairement à ce que l'on imagine souvent, le comportementalisme ne se réduit donc pas aux comportements directement observables de l'extérieur.

Le modèle étiologique du comportementalisme peut se prêter à une double lecture. La première lecture explique l'étiologie générale des comportements inadaptés. Celle-ci s'explique par l'histoire des conditionnement qui font formé le sujet, sa personnalité et ses comportements. Le conditionnement correspond à l'association entre un événement (stimulus) et une réponse du sujet (comportement). Certains conditionnements sont innés (le bruit violent entraîne de la peur chez le nourrisson), mais la grande majorité des conditionnements sont acquis tout au long de la vie et particulièrement durant l'enfance. Comme quasiment tous les êtres vivants, l'être humain a la capacité d'apprendre, c'est-à-dire de garder traces des relations qu'il entretient avec le monde : la mémorisation de ces relations (le feu brûle, j'éloigne ma main) est à la base des conditionnements. Ainsi, les troubles psychiques sont le résultat de l'interaction entre cette capacité automatique d'apprentissage et le hasard qui met en rapport stimulus (par exemple : ascenseur) et comportement (par exemple : crise d'angoisse) produisant une réaction inadaptée systématique (phobie des ascenseurs). Ainsi, ce qu'est un sujet, ce qu'il fait, ses réactions, dont ses comportements inadaptés, est la conséquence de l'ensemble des conditionnements qui le constituent et qui proviennent eux-mêmes de ses expériences de vie. La causalité théorique s'étend donc loin dans le passé du sujet.

Mais les modèles d'intervention thérapeutique proposent une deuxième lecture de ce modèle. Celle-ci consiste dans une causalité très ramassée : ce qui cause le comportement inadapté c'est un stimulus. Ce qui intéresse le thérapeute comportementaliste ce n'est pas l'histoire de la causalité du trouble, mais de bien repérer le stimulus déclencheur : ce qui cause la phobie, par exemple, c'est l'objet phobogène (et non l'histoire de son conditionnement). Le modèle d'intervention se réduit donc au schéma S (stimulus)-R (réponse). Le comportement inadapté est causé par le stimulus. La thérapie consiste donc à changer la réponse inadaptée. Dans l'exemple de la phobie sociale, la réponse (angoisse massive) est inadaptée au stimulus (des relations sociales). La thérapie, pour l'essentiel, consistera à exposer progressivement et

dans de bonnes conditions psychiques le sujet au stimulus phobogène (mettre le sujet en situation sociale réelle ou imaginaire en état de relaxation) de façon à ce que le conditionnement «relations-peur» laisse la place au conditionnement «relations-relaxation» incompatible avec le premier. Contrairement aux psychanalystes, les comportementalistes ne sont pas friands, dans leur approche thérapeutique, de la quête des causes lointaines. Leurs interventions sont concrètes et pragmatiques : ils agissent de façon active sur l'association S-R posant problème, en cherchant à la remplacer par une autre association moins inadaptée.

Là encore, j'ai réduit le modèle comportementaliste à ses fondements, il existe des apports qui le complexifie, comme celui de Skinner (conditionnement opérant), mais fondamentalement ces apports ne change pas le schéma de base S-R du comportementalisme. L'avantage d'un modèle aussi simple est d'être clair et testable (il l'a été, souvent avec succès). Par ailleurs, les interventions qui s'en revendiquent sont aussi en partie efficaces pour modifier les comportements problématiques. Cependant, ce modèle est aussi trop simple pour rendre compte de la complexité des troubles psychopathologiques, car il évacue des dimensions essentielles incluses dans ces problèmes, en particulier la dimension psychique. Dans le schéma S-R, il n'y a pas de place pour la vie psychique, ni comme stimulus, ni comme comportement. Tout se passe pour les comportementalistes comme si la vie psychique et ses caractéristiques (motivation, intention, interprétation, libre arbitre, capacité de choix, etc.) n'avait aucune part dans les troubles psychiques ni dans les comportements en général. Le sujet n'est jamais agent, il ne fait que réagir automatiquement aux *stimuli* qu'il reçoit du monde (y compris de son corps). L'estompage de la psyché est d'ailleurs pleinement cohérent avec l'histoire et l'intention du béhaviorisme de Watson, qui s'est constitué en réaction à un excès de subjectivisme et de méthode introspective et a voulu fonder une psychologie scientifique exclusivement sur des faits directement observables (les comportements). Naville, fervent défenseur du béhaviorisme en France dans les années d'après-guerre, se demande avec raison si le béhaviorisme est encore une psychologie (NAVILLE 1942). Autant la psychanalyse ramène à tout la causalité psychique (le déterminisme psychique), autant le comportementalisme des débuts s'efforce de l'évacuer totalement.

Mais, en psychopathologie, si l'on évacue la psychologie, on évacue aussi le sujet sensible et singulier, ce qui n'est pas sans risques éthiques.

Ainsi Naville peut exposer sans la moindre réticence éthique exprimée (en 1942, réédition de 1963) le cas célèbre du petit Albert, un an, chez qui des comportementalistes ont volontairement créé une phobie des rats blancs et finalement de plusieurs objets blancs, pour démontrer expérimentalement la causalité des phobies. Ce qui fut fait, mais Albert est reparti de l'hôpital avec un trouble psychique volontairement créé, ce qui aujourd'hui s'apparenterait à des mauvais traitements et relèverait du pénal. Naville avance aussi l'idée qu'il ne serait pas inacceptable (sauf pour les âmes sujettes à la sensiblerie, dit-il) de «supprimer» les «fous incurables», ni d'ailleurs d'ouvrir des camps de rééducation gérés par des comportementalistes pour accueillir, à vie s'il le faut, les inadaptés sociaux. Il n'est pas question d'affirmer que le comportementalisme conduisait systématiquement à une attitude déshumanisée sans autre éthique que la méthodologie scientifique (qui n'est pas une éthique) mais l'on peut se demander dans quelle mesure son modèle causal, parce qu'il évacuait le sujet sensible et singulier et le réduisait à une «machine» qui se comporte, pouvait restreindre les capacités empathiques des behavioristes. Trop d'objectivité et de «science» peut conduire, en psychologie appliquée, à occulter la subjectivité du patient qui est pourtant l'objet et la justification des exigences éthiques (ne pas nuire). Ainsi, les modèles de causalité psychopathologiques renvoient aussi à une ontologie, une conception de ce qu'est l'être humain : sexuel et inconscient pour les psychanalystes, sans agentivité ni vie psychique propre pour les comportementalistes. Dans les deux cas, le patient est soumis à des forces internes qui le dépassent et qui régulent malgré lui et automatiquement ses comportements.

Les dérives éthiques que j'évoque sont anciennes et sont marquées par une époque. Le fait d'accorder plus d'importance à son modèle qu'au bien-être de ses patients n'est pas le propre des comportementalistes historiques, l'histoire de la psychanalyse, de la médecine, en sont aussi marqué. Les ouvrages modernes comportementalistes qui évoquent ces dérives ne manquent pas d'en pointer les dérapages éthiques. Les comportementalistes d'aujourd'hui ne sont pas ceux d'hier et il n'y a pas de raison de douter maintenant de leur rigueur éthique et de leur sensibilité empathique. C'est pourquoi les attaques qui sont parfois portées contre les TCC d'aujourd'hui au nom des quelques exemples que je citais ne sont pas pertinentes. Elles le sont d'autant moins que les comportementalistes d'aujourd'hui fondent leur pratique sur des modèles qui ont

été considérablement complétés par l'apport des cognitivistes qui ont réintroduit la vie psychique (cognitions, émotions, etc.) dans les théorisations et les pratiques thérapeutiques. C'est pourquoi d'ailleurs l'on parle maintenant de thérapies comportementales et cognitives (TCC). Le schéma S-R a vieilli et a vu s'intercaler entre ses deux membres la notion d'organisme (O), recouvrant dans les modèles psychologiques la dimension psychique qui avait été évacuée par les comportementalistes historiques. La raison de cet enrichissement est simple : plusieurs études ont démontré qu'il était plus efficace, pour modifier les comportements, de prendre en compte la vie psychique du sujet (sa conscience, ses motivations, ses croyances). Ainsi, le comportement n'est plus conçu comme une simple réaction à un *stimulus*, mais comme le résultat de l'interprétation du *stimuli* par le sujet et des choix qu'il effectue en fonction de cette interprétation. De plus, ce modèle S-O-R est situé en interaction avec l'environnement (social, familial), ce qui implique, entre autre, que R puisse influencer par des effets rétroactifs sur S, introduisant une causalité circulaire. C'est en particulier avec Bandura que ces évolutions ont transformé les schémas comportementalistes traditionnels (BANDURA 1977).

3.2 ➔ Théories cognitives

Les sciences cognitives s'intéressent au traitement de l'information par le cerveau, et donc, aux fonctions psychiques concernées par ce traitement (mémoire, attention, jugement, conscience, processus inconscients, etc.). Elles élaborent des modèles psychopathologiques depuis les années 1970 (BANDURA 1977, BECK 1979). En clinique, l'approche cognitive a développé de nombreux modèles mettant l'accent sur les perturbations du traitement de l'information dans les troubles psychiques. Par exemple, les anxieux accroissent leur attention sur les signaux de danger et réduisent leur attention sur les signaux de sécurité. Ainsi le monde leur apparaît constamment très dangereux, produisant un état d'anxiété permanent. Autre exemple : certains schizophrènes seraient en difficulté pour sélectionner leur centre d'attention (attention sélective) et seraient parasités par les nombreux *stimuli* qu'ils ne parviennent pas à laisser de côté. Cela perturberait leur rapport au monde et aux autres. Outre ces multiples apports pour chaque trouble psychique, les approches cognitives ont développé la notion de «schémas» qui concerne l'ensemble des troubles psychiques et même au-delà, la vie psychique humaine en général. Ces schémas sont des constructions psychiques

intériorisées dans l'enfance, inconscientes, qui structurent et orientent la façon dont le sujet traite l'information qu'il reçoit du monde. Chaque être humain a donc en lui ces structures qui prédéterminent en partie son rapport au monde, ses pensées sur le monde et en conséquence ses émotions et son humeur. Les êtres humains en bonne santé mentale ont aussi ces schémas inconscients qui donnent un cadre nécessaire à l'interprétation du monde. Ces schémas «normaux» peuvent être positifs et faciliter leur rapport au monde, ou être négatifs mais peu actifs. Auquel cas ils participent de la sensibilité et de la personnalité du sujet sans lui poser de difficultés insurmontables. C'est par contre le cas des schémas dits «dysfonctionnels» qui sont négatifs et actifs et qui compliquent considérablement le rapport du sujet au monde. Ces schémas sont inconscients mais ils peuvent accéder à la conscience par un travail d'introspection. Plusieurs auteurs ont proposé des formulations directes de ces schémas dysfonctionnels (par exemple YOUNG 1990). Ce peut être : «Je dois tout le temps et toujours tout réussir», «Je dois tout le temps et toujours être aimé de tout le monde» (COTTRAUX 1998), «Je ne peux pas être aimé, cela n'a pas de sens que j'essaie de me lier aux gens», «Je ne vaud rien», «Je dois être prudent car les autres veulent me faire du mal» (CHAMBON 1997). Ces schémas, qui anticipent les effets des actions du sujet, placent celui-ci dans des impasses d'autant plus qu'ils tendent à se renforcer en s'autoconfirmant (le sujet ne perçoit que ce qui confirme ses schémas actifs), ils gênent son rapport au monde et produisent de la souffrance. Les schémas sont appris tout au long de la vie et particulièrement dans l'enfance.

Ce type de causalité globalement linéaire structure la théorie cognitive des schémas. Le modèle d'intervention thérapeutique, lui, prend en compte plusieurs éléments : d'une part un travail sur les schémas dysfonctionnels, d'autre part un travail sur leurs conséquences. Le travail sur les schémas consiste à amener le patient (par des entretiens, questionnaires, exercices divers) à prendre conscience de ses schémas dysfonctionnels. Mais cette prise de conscience, si elle est utile, ne suffit pas. Il convient ensuite d'amener le patient à remettre en cause les injonctions des schémas et de s'assurer que les relations aux autres et au monde peuvent être différentes de ce que le schéma affirme (non, le monde n'est pas si dangereux, les autres si méchants, le sujet si mauvais...). Il s'agit donc de lutter contre les effets pathologiques des schémas. Pour cela, suivant les troubles, de nombreux moyens sont

utilisables (exposition, jeux de rôle, développement de compétences spécifiques, affirmation de soi, relaxation, etc.), l'idéal étant que le travail thérapeutique modifie en profondeur le schéma, ou tout au moins le désactive. Ainsi la théorisation apparaît plutôt structurée par une causalité linéaire (expériences – schéma – troubles) et l'intervention introduit, elle, un modèle de causalité circulaire en mettant en œuvre un ensemble de rétroactions négatives pour contrer les effets des schémas.

Il est à noter que le modèle théorique des schémas est assez proche de celui de la psychanalyse évoqué plus haut : des structures inconscientes (relationnelles mêlant de près pensées et émotions) acquises durant l'enfance qui agissent sur la vie psychique actuelle. Mais il est tout aussi notable que les modèles thérapeutiques diffèrent grandement par le fait que les modèles d'interventions cognitives ne se contentent pas de la prise de conscience des schémas et de leurs aléas, mais luttent activement et concrètement contre leurs effets.

Rappelons enfin que les approches thérapeutiques cognitives s'intègrent aux approches comportementales et que toutes deux forment le champ des TCC. Ainsi, leurs différents modèles s'enrichissent mutuellement, donnant finalement des modèles de causalité complexes, interactifs, combinant linéarité et circularité.

4 ⇨ Les théories systémiques

Ces approches sont apparues dans les années 1960, avec les travaux de BATESON (1956), puis par exemple ceux de l'école de Palo Alto (WATZLAWICK 1967 ; pour une vue générale : EL KAÏM 1995). Elles s'appuient sur la théorie des systèmes, la cybernétique, le fonctionnalisme ethnologique et les théories de la communication. Les thérapies systémiques ont un positionnement théorique et technique original par rapport à la plupart des modèles thérapeutiques, y compris les deux courants présentés précédemment. D'une part, elles se fondent d'emblée sur un modèle de causalité circulaire (systémique), d'autre part, elles se décentrent de l'individu, pour s'intéresser aux systèmes au sein desquels il interagit. Tant la psychanalyse que les théorisations des TCC ont une vision individualiste du trouble (centrée sur le patient), ainsi qu'un modèle causaliste de base linéaire et donc évolutionniste : le trouble est le résultat d'une évolution, d'un enchaînement historique d'événements.

Les approches systémiques ne partagent aucune de ces deux conceptions. D'une part, le sujet n'y est qu'un élément constitutif des systèmes auxquels il appartient, et en particulier le système familial. Le «patient» des soignants d'autres approches devient le «patient désigné» pour les systémiciens, c'est-à-dire le patient qui est porteur des symptômes familiaux. Les individus sont des parties d'un tout et c'est ce tout qui compte. Les systémiciens s'intéressent à la totalité et non aux parties (aux individus). Plus précisément, ils s'intéressent à ce qui fait exister la totalité, c'est-à-dire aux relations entre les parties, à la communication. Les troubles psychiques participent des relations au sein du système, ils sont de la communication, ils sont donc une caractéristique du système et non un événement en eux-mêmes. Les troubles ont donc une fonction dans le système et ne sont pas la conséquence d'une évolution linéaire. En d'autres termes, il s'agit d'une conception fonctionnaliste du symptôme. La question n'est plus : «D'où vient le symptôme?», mais : «À quoi sert-il?» Qu'elle est sa fonction dans le système? En général, la fonction du symptôme est de maintenir l'équilibre dans le système par un ensemble de rétroactions favorisant la stabilité (quand du changement, nécessaire mais insupportable, menace le système). Le plus souvent, les troubles des uns, protègent les autres, c'est leur fonction. Il est vrai que l'on remarque régulièrement dans la pratique clinique que l'amélioration d'un patient peut être suivie de près par la détérioration de l'état d'un proche. C'est ainsi, par exemple, que le trouble psychique d'un enfant peut avoir pour fonction de sauvegarder le couple parental. Ou que l'état de dépendance familiale massive d'un patient a pour effet d'éviter la solitude dépressiogène d'autres membres de la famille. La perspective systémique éclaire bien ce type de situation.

Le modèle étiologique est donc systémique et fonctionnaliste. Le modèle d'intervention est très cohérent avec le modèle théorique étiologique. L'action thérapeutique consiste à favoriser les changements d'équilibre au sein du système afin que la fonction communicationnelle du trouble n'ait plus lieu d'être. Le dispositif particulier (vidéo, glace sans tain, etc.) de ces thérapies prend en compte la dimension systémique non seulement de la famille (entretiens familiaux) mais aussi de la thérapie (position «méta» des thérapeutes). Les interventions (recadrage, injonctions, paradoxes, «exercices», etc.) visent des modifications de la communication au sein du système et non le changement chez l'un

ou l'autre de ses membres. Tout cela donne une place particulière au problème de la causalité. Je cite WATZLAWICK (1967, p. 41):

Les causes possibles ou supposées d'un comportement n'ont qu'une importance secondaire, mais par contre l'effet de ce comportement dans l'interaction d'individus étroitement liés, devient un critère d'une importance primordiale.

Par rapport aux modèles présentés précédemment, il y a un changement de niveau: le trouble n'est plus la conséquence d'autre chose, mais il participe à la dynamique systémique des interactions qui elles-mêmes maintiennent le caractère total du système: tantôt effet, tantôt cause au sein de séquences de communication. L'attention est plus portée sur les effets du trouble sur le système que sur son origine. Il faut cependant noter que des dimensions plus individualistes et évolutionnistes ont pris place assez récemment dans les modèles systémiques, par le biais, par exemple de la notion de mythe familial (par exemple ANDOLFI 1995). Chaque membre du système, porterait en lui, du fait de ses expériences familiales précoces, un «mythe», ou un prototype plus ou moins conscient de ce que devrait être une «bonne famille». Les contradictions entre les «mythes familiaux» des membres du groupe familial pouvant être source de conflits et de troubles potentiels. Ainsi, les modèles systémiques, comme les autres modèles, ont intégré une double perspective causaliste: circulaire d'abord, linéaire ensuite.

5 ⇨ Faut-il viser à l'intégration de ces modèles ? Et à quelles conditions serait-ce possible ?

Du point de vue de la causalité, on voit de nettes différences entre les modèles d'intervention présentés: la psychanalyse se centre sur la cause première, originaire; le comportementalisme se contente de la cause (le *stimulus*) précédant le trouble; les systémiciens mettent l'accent sur l'effet du trouble et non plus sur sa cause, le trouble devient cause. De ce point de vue, on pourrait dire que le modèle psychanalytique est tourné vers le passé lointain des causes, le comportementalisme vers le passé immédiat et le présent, et les systémiciens vers le présent et le futur (la conséquence du trouble). Mais en réalité, on a vu presque à chaque fois qu'en évoluant au contact des patients et des pratiques psychothérapeutiques, chaque modèle théorique s'est complexifié et a débordé de son schéma de base vers des dimensions prises en compte par les autres modèles. Ainsi, la psychanalyse a développé des courants

comme l'*ego psychology* moins soucieux d'une quête infinie des causes et plus centrés sur le fonctionnement actuel du moi (c'est une psychanalyse plus psychologique que métapsychologique), ou encore des courants de thérapie familiale (ou institutionnelle) psychanalytique prenant en compte l'environnement réel actuel. Le comportementalisme s'est trouvé complété par le cognitivisme qui y a introduit le psychisme et son histoire et a de ce fait «agrandi» le champ de la causalité, y incluant aussi avec Bandura l'environnement actuel. Les systémiciens ont développé, par exemple, le concept de mythe familial qui réintroduit une dimension individuelle, subjective et historique dans leur modèle. Il semble que les modèles théoriques soient contraints à des évolutions du fait de leur contact avec la réalité clinique, et ces évolutions, sur certains points les rapprochent. Ainsi, l'expérience clinique impose certaines constantes qui ont contraint les modèles à évoluer pour les assimiler. Tous ces modèles intègrent, finalement, des dimensions causales linéaires remontant jusqu'à l'enfance du sujet, et circulaires prenant en compte les effets des troubles.

Cependant, face à ces différences dans les courants, plusieurs attitudes sont possibles. La première consiste à s'enfermer dans son courant d'élection, à le supposer totalisant (il explique tout) et à rejeter les autres courants. La seconde attitude repose sur la pluridisciplinarité : chaque courant, bien identifié dans ses actions et ses effets thérapeutiques, travaille en partenariat avec les tenants d'autres courants (comme des spécialités en médecine). La troisième attitude est intégrative et prône l'élaboration d'un métamodèle incluant chacun des courants considérés comme complémentaire des autres. C'est cette troisième attitude que je voudrais défendre.

Viser l'intégration, c'est supposer que les modèles présentés ici portent finalement sur le même objet, mais qu'ils l'envisagent sous trois angles différents. Ainsi se cantonner à un modèle, c'est se limiter dans la compréhension du cas clinique. Pour avoir une vision globale de la situation psychologique du sujet il convient donc d'intégrer dans le même temps plusieurs points de vue pertinents que sont la vie psychique profonde, les comportements et les relations actuelles. Le premier argument en faveur de cette position est que le sujet psychologique lui-même est «intégratif», dans le sens où ce qu'il est provient de l'interaction complexe et évolutive d'un ensemble d'éléments qui se manifestent très concrètement en situation clinique. Le sujet nous invite

donc à tenir compte de cette complexité et à ne pas le réduire à l'une ou l'autre de ses parties. Il faut rappeler le consensus de la communauté des psychologues sur la causalité multiple des troubles psychiques. À causalité multiple, intervention complexe. D'ailleurs, plusieurs études montrent que les thérapeutes, dans leurs pratiques, n'appliquent pas sans souplesse les injonctions de leur modèle, mais qu'ils s'adaptent aux besoins du patient et n'hésitent pas, souvent sans le savoir, à adopter des attitudes ou à avoir des actions qui relèvent d'autres modèles que le leur (KRAMER 2005). Ainsi, la pratique clinique impose de fait une attitude partiellement intégrative. Pourquoi ne pas en prendre acte et développer plus explicitement cette tendance ? Et cela d'autant plus que plusieurs études montrent que l'appartenance du thérapeute à tel ou tel modèle n'explique que très peu de la variance (2 à 10 % selon les études) de l'efficacité psychothérapeutique (DE ROTEN 2006). L'essentiel de cette variance étant expliqué par d'autres facteurs comme la personnalité du thérapeute, ses compétences relationnelles (empathie, écoute, compétences professionnelles), la personnalité du patient ou encore l'alliance thérapeutique (GALLO 2005, DESPLAND *et al.* 2000). Les principaux facteurs de changement se trouvent donc au-delà des modèles et il est pour le moins incongru de ce point de vue de revendiquer l'appartenance à tel ou tel modèle comme garant de l'efficacité thérapeutique.

Cela étant dit, l'intégration n'est pas sans poser plusieurs problèmes. Tout d'abord, la première question est : intégrer quoi ou si l'on préfère intégrer jusqu'où ? Notons que tous les modèles présentés ici sont déjà intégratifs car ils articulent ensemble plusieurs éléments de la vie psychique (par exemple les affects et la pensée, ou les comportements et les émotions, etc.) : aucun ne réduit plus la vie psychique à une seule dimension (comme le behaviorisme du début, par exemple). Le risque inverse nous est rappelé par l'adage : « tout est dans tout et réciproquement », la complexité est globale et à vouloir intégrer tout ce qui concerne le sujet il faudrait envisager l'intégration de tous les savoirs, ce qui n'aurait pas de sens scientifique ni d'utilité clinique. La question n'est donc pas d'intégrer ou non, mais d'intégrer quoi et pourquoi ? Il est nécessaire aussi de ne pas confondre les modèles théoriques avec les modèles d'interventions psychothérapeutiques qui sont apparus à plusieurs reprises différents au sein de chaque courant tant dans leurs constructions que dans leurs objectifs. Dernier problème, comment déterminer des critères de validation du métamodèle ?

Les raisons de l'intégration sont évoquées ci-dessus. Cependant, le but de l'intégration devrait être double : augmenter la compréhension psychologiques des troubles psychiques (leur étiologie, leur identification, leurs effets, leurs évolutions) et accroître l'efficacité psychothérapeutique. Ce double objectif détermine le périmètre de l'intégration : les différents courants des psychothérapies. Il s'agit donc d'envisager une intégration des modèles psychologiques (et non pas psychiatriques et biologiques), étant bien entendu que cela reste réducteur et éloigné de modèle intégratif plus vaste comme le modèle «biopsychosocial» d'Engels (ENGELS 1977). La cohérence de cette position tient dans le fait que les courants psychothérapeutiques ont les mêmes objectifs (soins psychiques) et le même objet d'étude (la vie psychique et ses différents niveaux : inconscient, conscient, comportement, relation) au point, comme on l'a vu, qu'ils se chevauchent en plusieurs endroits de leurs conceptions et de leurs pratiques. Il paraît donc justifié et utile de chercher à rendre compte globalement de cet «objet» qu'est la vie psychique et de lutter contre le morcellement de la psychologie. Cependant le double objectif évoqué plus haut nécessite une approche intégrative double : théorique et clinique. S'agit-il alors de cumuler tout ce qui existe en bricolant des raccords et en masquant les contradictions tant pour les conceptions théoriques que pour les modèles d'interventions ? Non. Il faut donc poser des critères de validation pour l'intégration. Les hypothèses théoriques devraient subir des validations scientifiques diverses (c'est-à-dire reposant sur des recherches empiriques et contrôlées, mêlant études clinique, corrélacionnelle, longitudinale, randomisée) et obtenir un certain niveau de preuve avant que d'être intégrées au modèle. Les modèles psychothérapeutiques devraient être testés aussi de façons diverses (report de cas unique, études empiriques de l'efficacité des interventions) avant que d'être reconnues comme utiles et intégrées au modèle d'intervention. Par «scientifique», il faut entendre une méthodologie qui respecte les principes de la démarche empirique (scepticisme initial sur les faits, rationalité, matérialisme méthodologique, réalisme, cf. LECOINTRE 2007). Les rapports entre théorie et clinique devraient être très étroits mais sans confusion, comme les rapports existant entre la recherche fondamentale et ses applications, ou entre la recherche biomédicale et la médecine. L'un vise la preuve, la solidité matérialiste du savoir, l'autre vise, en psychologie clinique, l'efficacité thérapeutique et celle-ci peut

reposer sur des facteurs encore scientifiquement incompris et d'apparence irrationnelle (comme l'effet placebo).

6 ➤ Conclusion

Tout savoir s'efforce pour l'essentiel de dégager des rapports de causalité, donnant ainsi sens au monde et guide pour agir dans le monde. La psychopathologie clinique, face aux événements psychiques déroutants que sont les troubles psychologiques, a produit depuis plus d'un siècle plusieurs modèles étiologiques théoriques structurant des stratégies d'intervention psychothérapeutiques. Des décennies d'expérience clinique ont relativement rapproché les conceptions de la causalité (des troubles, des améliorations) de chacun de ces courants. Malgré cela, persiste généralement en psychologie clinique une tendance au morcellement, à l'incompréhension et à des réflexes territoriaux, que ni les situations des patients, ni les données disponibles en psychologie et sur les psychothérapies ne justifient aujourd'hui. Pour dépasser ces antagonismes dont ne profitent ni la science psychologique ni les patients, il paraît pertinent de favoriser l'élaboration d'une psychologie clinique unifiée s'enrichissant des apports de chacun des courants qui aujourd'hui la composent, en s'appuyant sur une démarche associant exigence scientifique adaptée à son objet d'étude et évaluation empirique des effets de ses interventions. Pour cela, il convient de ne pas entretenir plus avant la confusion entre démarche scientifique de validation du savoir et démarche clinique (associant rationalité et intuition, savoirs et empathie, objectivité et subjectivité), confusion qui empêche la psychologie clinique de se constituer pleinement comme champ de savoir solide, cohérent et unifié.

➤ Références bibliographiques ◉

A

ANDOLFI M. (1995), « Famille/individu: un modèle trigénérationnel », in M. El Kaïm (dir.), *Panorama des thérapies familiales*, Seuil: 115-137.

B

BANDURA A. (1977), *Social learning theory*, Henglewood Cliffs, Prentice Hall.

BATESON G. (1980), « Vers une théorie de la schizophrénie », in G. Bateson, *Vers une écologie de l'esprit*, t. II, Seuil.

BECK A.T., RUSCH A.J., SHAW B.F. & EMERGY G. (1979), *Cognitive therapy of depression*, New York, Guilford Press.

BOURGEOIS M.-L. (2001), *Les Schizophrènes*, PUF.

BOTELLA C. et S. (2001), «De la recherche en psychanalyse», *Rev. Fr. Psychoan.*, n° spécial : «Courants de la psychanalyse contemporaine».

C

CHAMBON O. (1997), *Techniques de psychothérapie cognitive des psychoses chroniques*, Paris, Masson.

CONFÉRENCE DE CONSENSUS (1994), «Stratégies thérapeutiques à long terme dans les psychoses schizophréniques», 13 et 14 janvier 1994, Fédération française de psychiatrie.

CONFÉRENCE DE CONSENSUS (2003), «Schizophrénies débutantes, diagnostics et modalités thérapeutiques», janvier 2003, Fédération française de psychiatrie, <<http://psydoc-fr.broca.inserm.fr/conf&rm/conf/confschizo2/recommlongues.htm>>.

COTTRAUX J. (1998), *Les Thérapies comportementales et cognitives*, Masson.

D

DE ROTEN Y. (2006), «Ce qui agit effectivement en psychothérapie: facteurs communs ou agents spécifiques», *Bulletin de psychologie*, 59(6): 585-590.

DESPLAND J.-N., DE ROTEN Y., MARTINEZ E., PLANCHEREL A.-C. & SOLAI S. (2000), «L'alliance thérapeutique: un concept empirique», *Revue médicale suisse*, n° 685.

E

EL KAÏM M. (1995), *Panorama des thérapies familiales*, Seuil.

ENGELS G. (1977), "The need for a new medical model: a challenge for biomedicine", *Science*, 196: 129-135.

F

FREUD S. (1985), «Un cas de paranoïa en contradiction avec la théorie psychanalytique» [1915], in S. Freud, *Névrose, psychose et perversion*, PUF.

FREUD S. (1986), *La Naissance de la psychanalyse (1887-1902)*, PUF.

FREUD S. & BREUER J. (1985), *Études sur l'hystérie* [1892], PUF.

G

GALLO E., CERONI G.B., NERI C. & SCARDOVI A. (2005), "Fattori terapeutici specifici comuni (FTSC) nelle psicoterapie e negli altri trattamenti", *Rivista di Psichiatria*, 40(2): 63-81.

K

KRAMER U., DE ROTEN Y. & DESPLAND J.-N. (2005), «Les thérapeutes font-ils ce qu'ils disent faire? Comparaison entre prototypes idéaux et pratiques réelles pour plusieurs formes de psychothérapies», *Pratiques psychologiques*, 11: 359-370.

L

LAGACHE D. (1982), «La psychanalyse et la structure de la personnalité» [1960], in D. Lagache, *Œuvres complètes*, tome IV, *Agressivité et structure de la personnalité*, PUF.

LAZARTIGUES A. & LEMONNIER E. (2005), *Vivre et comprendre les troubles autistiques*, Ellipses.

LECOINTRE G. (2007), «Comprendre le matérialisme par son histoire», préface au livre de P. Charbonnat, *Histoire des philosophies matérialistes*, Syllepse.

N

NAVILLE P. (1963), *La Psychologie du comportement* [1942], Gallimard.

R

REICH W. (1999), *Irruption de la morale sexuelle*, Payot.

V

VGOTSKI L.S. (1999), *La Signification historique de la crise en psychologie* [1927], Lausanne, Delachaux et Niestlé.

W

WATZLAWICK P., HELMICK BEAVIN J., JACKSON D.D. (1972), *Une logique de la communication* [1967], Seuil.

WESTEN D. (1999), "The scientific status of unconscious processes. Is Freud really dead?", <www.psychomedia.it/rapaport-klein/westen99.htm>.

Y

YOUNG J. (1990), "Cognitive therapy for personality disorders : a schema focused approach", Sarasota, Professional Resource Exchange.

☞ Notice biographique ☞

Cyrille Bouvet est maître de conférences de psychologie clinique, Université Paris Ouest Nanterre la Défense, UFR SPSE, laboratoire EVACLIPSY (200, avenue de la République, 92001 Nanterre. <cbouvet@u-paris10.fr>). D'abord psychologue clinicien en hôpital de jour psychiatrique durant une quinzaine d'années, Cyrille Bouvet est depuis cinq ans enseignant-chercheur en psychologie clinique. Spécialisé dans la réhabilitation psychosociale en psychiatrie et les troubles psychiques handicapants (comme la schizophrénie), il publie des articles de réflexions et de recherches dans des revues scientifiques et professionnelles.

Avertissement. Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur et des auteurs. Les textes et illustrations figurant sur ce site peuvent être consultés et reproduits sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage personnel, scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document selon ce modèle :

Cyrille Bouvet, «Causalité(s) en psychopathologie: des modèles inconciliables?», in Gérard Lambert et Marc Silberstein (dir.), Matière première. Revue d'épistémologie [en ligne]. Nouvelle série, N° 1/2010 : Épistémologie de la médecine et de la santé, Éditions Matériologiques. Mis en ligne le 7 octobre 2010. URL: www.materiologiques.com

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.



Catherine Dekeuwer **Définir** **la maladie génétique**

J'aimerais apporter une contribution à la réflexion philosophique sur la définition de la maladie génétique¹ en montrant que trois critères sont nécessaires, ensemble, pour considérer une maladie comme génétique. Avant d'entrer dans le vif du sujet, deux remarques s'imposent. D'abord, la notion de maladie est en elle-même difficile à définir²; doit-on pour autant considérer cette difficulté comme un obstacle rédhibitoire à toute tentative de définition de certaines maladies comme génétiques? Je pense au contraire que les analyses présentées ici pourraient alimenter les réflexions sur ce qu'est, plus généralement, une maladie. Ensuite se pose la question de savoir d'où partir pour définir les maladies génétiques. La notion de maladie génétique renvoie à l'idée qu'un ou plusieurs gènes en sont des causes; toute la question est alors de savoir comment ces causes sont conçues. En outre, le développement d'une pathologie ne peut pas être attribuée aux seuls gènes. Le problème est donc de savoir pourquoi, pour certaines maladies, les causes génétiques sont privilégiées par rapport aux causes environnementales. Pour résoudre ce problème, faut-il alors s'appuyer sur une recension des usages actuels de l'expression «maladie génétique» ou sur une analyse conceptuelle de la causalité génétique à partir de laquelle une frontière pourrait être clairement tracée entre maladies génétiques et non généti-

1. HULL (1979), HESSLOW (1984), STERELNY & KITCHER (1988), GIFFORD (1990), GANNETT (1999), SOBER (2001), KITCHER (2003), MAGNUS (2004).

2. Cf. LAMBERT (2009). (NdÉ.)

ques? Quelle place doit être faite à l'histoire de la médecine et de la biologie dans cette recherche? Il me semble que les usages de l'expression «maladie génétique» renvoient à la fois à une série de problèmes que les médecins ont à résoudre, et en même temps aux concepts élaborés à cette fin. Comprendre la notion de maladie génétique demande donc d'emprunter une voie à la fois conceptuelle et historique.

Un article de David MAGNUS (2004) intitulé «The Concept of Genetic Disease» me sert de point d'appui pour présenter trois approches de la causalité génétique. Magnus distingue en effet trois concepts concurrents de la maladie génétique, qui correspondent à trois approches de la causalité. Selon lui, aucun n'est véritablement satisfaisant, ce qui l'amène à considérer les usages du concept pour comprendre pourquoi la causalité génétique est souvent privilégiée dans l'étude et la prise en charge de certaines pathologies. Cependant, alors que Magnus dénonce les insuffisances de chaque approche et réduit le sens de la maladie génétique à ses usages médicaux, je défendrai ici l'idée que ces trois conceptions de la causalité, une fois précisées, fonctionnent *ensemble* dans la définition de certaines maladies comme génétiques.

1 ⇨ Des problèmes concernant la causalité génétique

Magnus distingue donc trois approches de la causalité génétique, chacune posant des problèmes spécifiques : il est donc impossible, selon lui, de s'entendre sur une définition du concept de maladie génétique en l'absence de tout fondement épistémologique ferme. Je propose, dans cette première partie, de reprendre ces trois approches et d'en discuter les critiques.

Premièrement, une maladie est génétique si elle découle de l'action causale directe d'un ou de plusieurs gènes. Pour expliquer la notion d'action causale directe, Magnus se réfère à un article de Fred GIFFORD (1990, p. 329) dans lequel il affirme que pour être considéré comme génétique, «le trait doit être l'effet *spécifique* d'une cause génétique, c'est-à-dire que le trait doit être décrit ou individué de façon à correspondre adéquatement à ce que le gène cause spécifiquement». Magnus ne prend même pas la peine de préciser sa critique de cette conception de la causalité : selon lui, le développement d'une maladie est tellement complexe qu'il est vain d'en chercher des causes génétiques directes. Pourtant, il confond ici causalité directe et causalité spécifique. Gifford ne parle pas de causalité directe ; il utilise le concept d'effet spécifique

défini comme une correspondance établie par les chercheurs entre, d'une part, ce que le gène cause spécifiquement et, d'autre part, le phénotype tel qu'il est décrit et individué. L'effet est spécifique si la modification d'un gène a des effets sur le trait considéré mais non sur d'autres traits. Le phénotype doit être individué, c'est-à-dire correspondre à une unité de description précise, ni trop large ni trop étroite. Gifford prend d'abord l'exemple de l'aptitude à apprendre une langue. S'il est admis que les gènes ont un effet sur l'aptitude à apprendre une langue, la capacité d'apprendre *le français* n'est pas pour autant considérée comme génétique: le trait est alors individué trop spécifiquement. Deuxième exemple: l'hypercholestérolémie ne peut pas être considérée comme une maladie génétique. Il faut en effet distinguer les formes familiales et sporadiques de cette maladie, en individuant ainsi le phénotype rapporté à une cause génétique. Cette perspective a l'avantage d'attirer notre attention sur la nécessité de bien identifier et délimiter les phénotypes pour lesquels des causes génétiques pourraient être recherchées. Ainsi, l'absence de description et d'individuation claire de la schizophrénie a rendu difficile la recherche de ses déterminants génétiques (MAZIADE *et al.* 2003).

Le concept d'effet spécifique introduit les problématiques de la pathologie moléculaire. Gifford explique qu'il faudrait accorder moins d'importance à la question de savoir si un trait est génétique, mais se demander plutôt «quelles sont les étapes impliquées dans les voies de biosynthèse ou de développement» (GIFFORD 1990, p. 328). Or, «la pathologie moléculaire cherche à expliquer pourquoi une modification génétique donnée devrait conduire à un phénotype clinique particulier. [Elle] nous demande de comprendre l'effet d'une mutation sur la quantité ou la fonction du produit du gène, et d'expliquer pourquoi la modification est ou non pathogénique pour toute cellule, tissu ou étape de développement particuliers» (STRACHAN & READ 2004, p. 418). La pathologie moléculaire cherche à répondre à ce type de question: «Pourquoi est-ce que la perte de fonction de la protéine FMR1, impliquée dans le transport de l'ARN du noyau vers le cytoplasme causerait un retard mental et un macro-orchidisme (syndrome de l'X fragile)?» (*Ibid.*, p. 416.) Dans cette perspective, un gène serait donc une cause d'une maladie en ce sens que l'identification d'une séquence d'ADN permettrait d'expliquer spécifiquement les différentes étapes du processus qui conduit un individu à être malade. Cette approche de la causalité (une fois bien comprise) renvoie donc à l'influence spécifique des gènes

sur la manière dont les organismes se développent. Elle relève d'une démarche explicative relative à l'individu biologique : il s'agit de retracer une histoire causale relative aux effets spécifiques des produits d'un ou de plusieurs gènes.

La seconde approche de la causalité pertinente dans le cadre d'une définition de la maladie génétique n'est plus relative à l'individu, mais à des populations. Une maladie est génétique si « dans une population donnée la covariance entre ce trait et certains facteurs génétiques est plus grande que la covariance entre ce trait et d'autres facteurs (non-génétiques) » (MAGNUS 2004, p. 235). Magnus reprend encore un critère à Gifford : « Un trait est génétique (relativement à une population P) si ce sont les facteurs génétiques qui “font la différence” entre les individus qui ont le trait et le reste de la population P » (GIFFORD 1990, p. 333). Ce concept de cause est statistique et n'a de sens que relativement à des populations. L'analyse de la variance est en effet utilisée pour mesurer la contribution causale des facteurs génétiques et environnementaux dans une population donnée. Selon Magnus, cette approche n'est pas non plus convaincante, justement parce qu'elle est relative à une population : en fonction de la population étudiée, un facteur peut être considéré comme génétique ou non. L'identification d'un facteur génétique d'une maladie dépend donc de la population sélectionnée pour l'étude, ce qui fait du concept de maladie génétique un concept trop relatif. Magnus fait une analyse proche de celle proposée par HESSLOW (1984) et prend l'exemple d'un puits contaminé par un pathogène dans un village. La moitié seulement des villageois tombe malade et les chercheurs supposent alors que leurs gènes leur confèrent une résistance au pathogène. À l'échelle du village, la covariance des facteurs génétiques et de la maladie est proche de 1, tandis que la covariance de la maladie et du pathogène est proche de 0,5. La maladie est donc, relativement à la population du village, génétique. Par contre, à l'échelle de la planète, la maladie serait probablement environnementale. Pour Magnus, ce type d'approche statistique de la causalité n'est pas une bonne approche pour définir certaines maladies comme génétiques. J'ajouterai que la mise en évidence d'un facteur de risque n'est précisément pas l'identification d'une cause³.

3. Cf. l'article d'Élodie Giroux, « Facteurs de risque et causalité en épidémiologie », dans ce volume. (NdÉ.)

Il faut préciser cette critique et distinguer plusieurs approches statistiques des causes génétiques. Magnus fait référence à une maladie à laquelle les villageois seraient plus ou moins susceptibles. Ce sont les *études d'association* qui permettent d'identifier les déterminants du *terrain génétique*, mais non des maladies génétiques. La notion de terrain est en effet pertinente quand on se demande quels sont les déterminants génétiques qui expliquent la résistance de certains individus à des maladies dont l'étiologie est environnementale, virale par exemple. Les études d'association sont alors adaptées: elles cherchent à corrélérer une différence de fréquence allélique pour un même locus entre des sujets non apparentés atteints ou non atteints par la pathologie. Un allèle est alors «associé» à une maladie s'il est présent plus fréquemment parmi les malades que parmi les témoins. Il confère un risque accru de maladie par rapport à la population générale, exprimé par le risque relatif. Le choix des populations de référence s'avère alors certes crucial, mais ce n'est pas, à mon sens, le principal problème de ces études d'association. On cherche en effet ici à expliquer pourquoi certains individus sont plus susceptibles que d'autres d'attraper des maladies virales ou des cancers par exemple. Des facteurs génétiques de susceptibilité sont alors mis en évidence, mais on aurait beaucoup de mal à justifier que ces maladies soient considérées comme des maladies génétiques. Il serait difficile de défendre par exemple le Sida comme maladie génétique, même si des cas de résistance au virus HIV ont été répertoriés et même si des facteurs génétiques de résistance étaient identifiés. L'erreur est donc ici dans la confusion du terrain génétique et des maladies génétiques.

Les études d'association sont surtout pertinentes pour l'étude des maladies multifactorielles à hérédité complexe telles que l'asthme, le diabète ou certaines formes de la maladie d'Alzheimer⁴. Pour ces maladies, une agrégation familiale est souvent observée, mais le mode de transmission reste souvent inconnu. Il faut donc trouver une méthode pour étudier les déterminants génétiques de ces maladies sans connaître leurs modes de transmission, et l'étude d'association répond à cette exigence. En 1993, des études d'association ont ainsi montré une corrélation statis-

4. Ces maladies multifactorielles sont aussi le plus souvent multigéniques: un grand nombre de gène semble impliqué dans le développement de ces pathologies. Cf. CAMPION (2001), STRACHAN & READ (2004) et FEINGOLD (2005).

tique forte entre un polymorphisme du gène APOE (apolipoprotéine E), nommé e4, et les formes les plus fréquentes de la maladie d'Alzheimer, qui sont des formes non mendéliennes. Cette corrélation entre l'allèle e4 du gène APOE et la forme sporadique de la maladie d'Alzheimer a, depuis, été confirmée par des études indépendantes; elle est valide au moins pour les populations qui descendent d'ancêtres communs européens. On a calculé que la maladie d'Alzheimer est deux à trois fois plus fréquente dans le groupe des individus hétérozygotes qui portent un allèle e4 que dans la population générale. Elle est 9 à 15 fois plus fréquente pour les homozygotes, qui portent cet allèle en deux exemplaires. On parle alors du gène APOE comme d'un gène de susceptibilité pour les formes sporadiques de la maladie d'Alzheimer: plus de la moitié des personnes atteintes de maladie d'Alzheimer portent l'allèle e4, mais de nombreuses personnes qui portent cet allèle ne développeront jamais la maladie. L'absence de cet allèle chez une personne, enfin, ne signifie pas qu'elle ne sera pas atteinte de maladie d'Alzheimer. Cependant, les associations sont souvent faibles, les études peu reproductibles et on peine à saisir le sens physiopathologique d'une association entre un allèle qui sert de marqueur et une fréquence accrue de maladie. Le facteur de risque reste en effet souvent sans signification biologique, ce qui rend difficile, voire impossible, son interprétation comme une cause génétique. Par exemple, on cherche aujourd'hui à associer des marqueurs polymorphes appelés SNP (*Single Nucleotide Polymorphism*) à des pathologies complexes, mais on sait d'avance que ces marqueurs n'ont pas de fonction biologique. S'ils n'interviennent pas sur l'apparition des pathologies, il est alors difficile de ranger les maladies multifactorielles dans la catégorie des maladies génétiques.

Les *études de liaison génétique*, enfin, concernent surtout les maladies à hérédité mendélienne pour lesquelles des altérations d'un gène ont un effet majeur et spécifique sur le phénotype (maladie de Huntington, mucoviscidose et cancer du sein et des ovaires dit héréditaire, par exemple). Elles reposent sur la coségrégation d'un phénotype et de certains allèles d'un gène dans des familles. Pour une maladie monogénique comme la maladie de Huntington, par exemple, des marqueurs génétiques ont été identifiés grâce à des études de liaison. Ils ont permis de prédire quels individus, dans une famille, risquaient d'être malades; des tests génétiques sont réalisés dès 1986. À cette époque pourtant, la protéine codée par le gène proche du marqueur et les mécanismes moléculaires de la maladie ne sont pas identifiés. Le «gène qui cause la mala-

die» n'est donc pas identifié par l'analyse de liaison, mais seulement un marqueur. C'est en 1993 qu'un gène est identifié et que les recherches sur la protéine huntingtine peuvent commencer. Récapitulons. Les analyses de liaison permettent de localiser les gènes impliqués dans l'apparition des maladies monogéniques dont le mode de transmission est connu. L'analyse statistique, dans ce cas, repose sur la connaissance du mode de transmission de la maladie, ce qui donne sens à la notion de maladie «génétique». Pour autant, comme dans le cas des études d'association, tant que la fonction du gène n'est pas connue, il est difficile d'extrapoler d'une liaison à une relation de causalité. Tous ces exemples montrent donc que l'approche statistique ne donne pas en elle-même de moyen de comprendre les processus pathologiques, et qu'elle ne permet parfois que d'identifier des facteurs de risque. Elle conduit, par contre, à mettre au point des outils qui permettent d'identifier des personnes prédisposées à certaines maladies. Elle souligne, enfin, l'importance de l'aspect héréditaire des maladies considérées comme génétiques.

Selon Magnus, l'exemple de la contamination du puits montre qu'intuitivement nous considérons que l'agent pathogène est la cause de la maladie parce que nous estimons pouvoir agir sur cette cause, par exemple en nettoyant l'eau. Selon cette troisième approche de la causalité, ce ne sont plus des arguments scientifiques mais des intérêts concernant la prise en charge médicale qui expliquent pourquoi certaines maladies sont considérées comme génétiques. Dans cette perspective, une maladie est donc génétique quand ses déterminants génétiques paraissent être plus faciles à manipuler que ses déterminants environnementaux pour la prévenir ou la traiter. Magnus rappelle cependant qu'il n'existe pas de thérapie génétique ; la justification thérapeutique de cette approche instrumentale est donc largement compromise. De plus, certaines maladies réputées génétiques sont prises en charge médicalement par des modifications environnementales. Un régime alimentaire, par exemple, suffit à empêcher l'apparition des symptômes caractéristiques de la phénylcétonurie. Finalement, faut-il réduire cette approche instrumentale à sa dimension préventive ? Les maladies seraient considérées cette fois comme génétiques lorsqu'un test génétique serait mis sur le marché et permettrait de prédire qui risque ou non d'être malade. Cette solution montre bien les limites actuelles de la prise en charge de la grande majorité des maladies dites génétiques : le plus souvent, un test génétique sert à prévenir une maladie en sélectionnant les naissances.

Magnus considère ces conceptions de la causalité l'une après l'autre et les rejette. Mais pourquoi ne pas considérer que ces trois approches de la causalité contribuent, ensemble, à définir une maladie comme génétique? Ces concepts ne me semblent pas «antagonistes» (*competing*) mais ils font référence à trois aspects de la médecine. D'abord, la pratique médicale : les recherches en génétique permettent effectivement des prises en charges médicales spécifiques qui concernent des individus et des familles. Ensuite, la recherche d'une meilleure compréhension des maladies, qui comprend la localisation des déterminants génétiques des maladies ainsi que l'explication de leur rôle physiopathologique.

Pour mieux comprendre pourquoi il est difficile de définir un concept de maladie génétique, une distinction importante doit enfin être introduite. Lenny Moss (2004) distingue deux concepts de gène : le gène-P (pour prédiction) et le gène-D (pour développement). Le premier concept apparaît dans le contexte des études de liaison qui permettent d'identifier des marqueurs puis des «gènes de maladie». Dans ce cas, en effet, la liaison entre le gène et la maladie est suffisamment forte, même si rien n'est connu sur le mode d'action spécifique du gène, pour que des prédictions soient possibles. Ce concept a donc une valeur essentiellement instrumentale : le gène est défini par sa relation à un phénotype, mais connaître la séquence d'ADN ou la manière dont le gène agit n'est pas nécessaire à la prédiction. Par contre, la prédiction permet la prévention. Le gène-D est défini comme une ressource pour le développement, en elle-même indéterminée par rapport au phénotype. Dans ce dernier cas, la séquence d'ADN est connue et considérée simplement comme un élément des processus de développement, sans référence à un phénotype particulier. Un même gène peut être considéré comme gène-P ou gène-D. Par exemple, certaines versions du gène BRCA1 sont corrélées à un risque important de cancer du sein et des ovaires. Dans ce cas, on peut considérer BRCA1 comme un gène-P. Mais quand le gène BRCA1 n'est pas considéré comme «le gène du cancer du sein», mais comme un modèle pour la synthèse de protéines présentes dans plusieurs cellules et plusieurs tissus, alors c'est un gène-D. Ces protéines peuvent être étudiées dans chaque contexte cellulaire et tissulaire sans prendre en compte le phénotype (cancer du sein). Or, si le gène-P est un outil de prédiction utilisé afin d'obtenir un bénéfice médical ou économique, le gène-D a plutôt une valeur explicative.

Armés de ces distinctions conceptuelles, il convient maintenant de développer des exemples qui permettent d'étayer la thèse d'un fonctionnement commun de ces trois approches de la causalité pour définir une maladie comme génétique.

2 ⇨ Le concept de maladie moléculaire

Le concept de maladie génétique doit être compris dans le cadre des recherches en pathologie moléculaire. J'aimerais ici préciser ce concept et montrer également comment il s'articule aux recherches médicales sur la transmission héréditaire des maladies. Magnus cite les travaux de Linus Pauling et rappelle l'importance du concept de «maladie moléculaire» qui apparaît dans son célèbre article de 1949 (PAULING *et al.* 1949, cf. FELDMAN & TAUBER 1997). Mais il ne souligne pas assez la spécificité de l'approche de Pauling et ses liens avec les travaux du généticien James NEEL (1949).

Pour bien comprendre le concept de maladie génétique qui est ici en jeu, il est utile de revenir à une question simple. Pourquoi, alors que certains individus ne ressentent les crises drépanocytaires qu'à une haute altitude, la drépanocytose est-elle considérée comme une maladie génétique et non une maladie environnementale? Un des éléments de réponse réside dans l'analyse moléculaire de cette maladie effectuée dans les années 1950 (PAULING *et al.* 1949, INGRAM 1957, INGRAM & STRETTON 1959). L'équipe de Pauling rappelle, au début de l'article de 1949, que la drépanocytose se caractérise par une anémie sévère, qui résulte d'anomalies cellulaires. Le trait anémique est une forme moins sévère de la maladie, qui n'est pas ressentie dans la plupart des cas par les individus. Cette maladie est définie, au niveau cellulaire, par deux caractéristiques: la forme spéciale que prennent les globules rouges, dite «en faucille», et leur rigidité. À l'inverse, les globules rouges des individus qui ne sont pas malades sont flexibles et ont une forme de disque concave sur les deux faces. Or, deux hypothèses peuvent expliquer le processus par lequel les globules rouges prennent une forme en faucille. La première est une hypothèse cellulaire: la rigidité des globules rouges s'expliquerait par les propriétés de la membrane cellulaire. La seconde est moléculaire: ce processus serait lié aux propriétés chimiques et physiques de l'hémoglobine, une protéine présente dans ces cellules. L'équipe de Pauling choisit l'hypothèse moléculaire et étudie ces propriétés chimiques en comparant les hémoglobines des individus

anémiques, des individus qui ont un trait anémique, et de ceux qui n'ont aucune anomalie des globules rouges. La notion de maladie moléculaire vient donc d'une part de la distinction de trois niveaux d'analyse, les niveaux clinique, cellulaire, et moléculaire. Elle repose d'autre part sur le choix d'une hypothèse explicative de niveau moléculaire. L'ADN est également une molécule, et il est alors assez facile de compléter ce schéma par la référence à un niveau encore plus élémentaire : celui des gènes qui codent les protéines.

Mais ce n'est pas tout. Pour comprendre ce que signifie une cause moléculaire, il faut être attentif au fait que les chercheurs essaient de trouver la *meilleure* explication du phénomène caractéristique de la drépanocytose. Or, l'explication la meilleure est moléculaire : il est possible de distinguer deux types d'hémoglobine, qui diffèrent par leur charge électrique, la différence de forme des cellules étant finalement rapportée à cette propriété biochimique. Arrêtons-nous un instant sur l'explication causale telle qu'on peut la lire dans cet article. L'hémoglobine, dont la fonction est de transporter l'oxygène dans l'organisme, est décrite en termes biochimiques, comme une succession d'acide aminés (dont certains sont chargés électriquement). Or, en raison de la charge électrique de la protéine, chaque molécule d'hémoglobine anormale comporterait une région complémentaire d'une région d'une autre molécule d'hémoglobine ; ces molécules pourraient donc s'associer. L'association des molécules d'hémoglobine conduirait à leur alignement dans la cellule ; cette dernière prendrait alors une forme en faucille. L'explication moléculaire est à la fois spécifique et complète.

Du concept de maladie moléculaire à celui de maladie génétique, il n'y a alors qu'un pas. En 1957, Ingram analyse les acides aminés qui composent les deux types de protéines et trouve l'acide aminé modifié. Il explique :

La séquence de paires de bases le long de la chaîne d'acide nucléique fournit l'information qui détermine la séquence des acides aminés dans la chaîne polypeptidique de laquelle le gène particulier, ou le fragment d'acide nucléique, est responsable. Une substitution dans les acides nucléiques conduit à une substitution dans le polypeptide. (INGRAM 1957, p. 328.)

Si l'on considère que les termes «conduire» et «être responsable» expriment une relation de causalité, on comprend pourquoi la séquence d'ADN est considérée par les chercheurs comme la cause première de la

maladie. Dans le cas où un organisme est l'objet d'étude, il est en effet impossible de remonter en deçà de la séquence d'ADN pour expliquer la présence de la protéine anormale.

L'exemple de la découverte du «gène de la mucoviscidose» explique aussi en quel sens une maladie est considérée comme génétique. En 1985, les principaux symptômes de cette maladie sont connus : des affections chroniques des poumons et des voies respiratoires et, dans certains cas, une insuffisance d'enzymes pancréatiques. Le diagnostic de cette maladie est effectué en mesurant en particulier le taux de chlore dans la sueur. Depuis 1983, les chercheurs savent que les tissus des malades ne sont pas bien perméables au chlore ; la maladie serait peut-être causée par un défaut dans le transport de cet ion. Mais ils ne comprennent pas «la cause» de ces symptômes et cherchent un «défaut fondamental» responsable de la maladie :

En dépit d'efforts de recherche de grande ampleur, le défaut fondamental de la mucoviscidose reste non identifié. Par conséquent, les anomalies métaboliques, sur lesquelles les études biochimiques sont basées, sont probablement des conséquences secondaires ou tertiaires du défaut premier. (Tsui *et al.* 1985, p. 1054.)

La stratégie directe d'étude de la protéine anormale employée par Pauling est ici impossible : le défaut primaire explicatif de la pathologie est inconnu. C'est alors la méthode de génétique inverse qui permet aux chercheurs d'identifier la cause de la mucoviscidose. Philip Kitcher décrit ainsi cette stratégie de recherche :

Même si les chercheurs en biomédecine sont initialement totalement ignorants des processus physiologiques qui vont de travers dans une maladie donnée, en sachant comment la maladie est transmise dans un échantillon suffisamment grand de familles, ils peuvent parfois isoler le locus qui en est responsable. La stratégie est de trouver des marqueurs génétiques [...] associés à la transmission de la maladie, de limiter le locus à une région particulière du chromosome [...], de choisir des gènes candidats, et, enfin, de cloner et séquencer le gène cherché. La connaissance du gène peut alors permettre de comprendre suffisamment la protéine pour éclairer la base causale de la maladie. (KITCHER 1994, p. 522.)

Dans le cas de la mucoviscidose, le gène responsable de la pathologie est identifié et séquencé en 1989 : il code la protéine CFTR (*Cystic Fibrosis Transmembrane Conductance Regulator*), impliquée dans le transport des ions.

La notion de gène candidat est importante : à partir d'une étude de liaison, une région de l'ADN est localisée. Le séquençage fournit aux chercheurs une longue liste de bases azotées, dans laquelle ils peuvent repérer plusieurs groupes de bases azotées qui correspondent à des gènes. Le problème est alors de savoir lequel, parmi tous ces gènes possibles, pourrait être «le gène de la mucoviscidose». Or, à partir de la séquence d'ADN, il est possible de déterminer la séquence des acides aminés de la protéine, et, dans certains cas, sa fonction. Un ordre de priorité est alors défini par les chercheurs en fonction des structures et des fonctions des protéines codées par les gènes candidats. Pour certains de ces gènes, il est plausible que leur modification cause les symptômes caractéristiques de la maladie. L'ordre de priorité est ainsi défini en fonction de données qui ne sont pas génétiques mais physiologiques et pathologiques.

Le gène codant la protéine CFTR est sélectionné de la manière suivante (RIORDAN *et al.* 1989) : on sait que la conductance de l'ion chlore à travers la membrane des cellules diminue chez les individus malades ; un bon gène candidat serait donc une séquence d'ADN qui code une protéine contribuant à former un canal ionique ou une séquence d'ADN codant une protéine impliquée dans la régulation des canaux ioniques. Deuxièmement, on identifie l'ARN correspondant à l'ADN de CFTR dans les tissus des poumons, du côlon et des glandes sudoripares ; cette localisation est cohérente avec les symptômes caractéristiques de la pathologie. Enfin, CFTR est analysée : deux ensembles d'acides aminés forment un domaine capable de traverser la membrane plasmique d'une cellule ; cette donnée est cohérente avec l'hypothèse selon laquelle la maladie résulterait du dysfonctionnement d'un canal ionique. De plus, certains acides aminés de cette protéine sont susceptibles de former des domaines de liaison pour des molécules intervenant dans les processus de régulation ; cette donnée est également cohérente avec l'hypothèse d'un rôle régulateur de la protéine recherchée. Enfin, l'ordre des acides aminés de cette protéine est comparé à celui de protéines d'autres espèces dont la fonction est connue. Cette comparaison permet de conclure que la protéine codée par le gène candidat est vraisemblablement impliquée dans le transport des ions de part et d'autre de la membrane plasmique des cellules. C'est donc la cohérence de ces données physiopathologique qui donnent son sens au concept de maladie génétique. N'oublions pas enfin cette remarque : l'étude de génétique

inverse suit une étude de liaison génétique et n'est donc possible que si le mode de transmission de la maladie est connu. Qu'est-ce qu'une maladie génétique alors? C'est une maladie héréditaire dont le défaut primaire (moléculaire) a été identifié et permet d'expliquer l'ensemble des caractéristiques de niveau supérieur de la pathologie.

L'importance de la dimension héréditaire d'une maladie dans sa définition comme maladie génétique peut également être soulignée dans l'exemple de la drépanocytose. Ingram affirme :

[L'hémoglobine] est une protéine anormale qui est transmise d'une manière strictement mendélienne; il est possible maintenant de montrer, pour la première fois, que l'effet d'une mutation dans un seul gène est un changement dans un acide aminé de la chaîne polypeptidique de l'hémoglobine de la fabrication de laquelle ce gène est responsable. (INGRAM 1957, p. 326.)

C'est précisément cette identité entre le gène entendu comme facteur héréditaire et le gène compris comme séquence d'ADN qui explique l'importance de ces travaux dans l'élaboration du concept de maladie génétique. Jusqu'en 1949, la drépanocytose est considérée comme une maladie à transmission dominante et à expressivité variable (l'expressivité variable renvoie au fait que les globules rouges en forme de disque prennent la forme de faucilles uniquement quand la pression en oxygène diminue). Les généticiens supposent à cette époque qu'un seul exemplaire du gène muté est suffisant pour expliquer la fréquence des individus plus ou moins anémiques dans les familles; le même allèle du gène s'exprime de plusieurs manières différentes. Neel, dans un article de 1949 (NEEL 1949), montre que la drépanocytose se transmet de manière récessive. Ses prédécesseurs se trompent parce que l'entité clinique n'est pas bien définie: le trait anémique et l'anémie falciforme ne sont pas suffisamment distingués. Neel distingue les phénotypes (trait anémique et anémie falciforme), dresse des arbres généalogiques et observe la fréquence des individus atteints et non atteints du trait anémique et de l'anémie falciforme. Ses calculs montrent que l'hypothèse d'une transmission récessive doit être préférée. Pauling explique (PAULING *et al.* 1949) qu'il est arrivé à cette même conclusion avant la publication de l'article de Neel. L'expérience de biologie moléculaire de 1949 le conduit en effet à penser que le mode de transmission de la maladie est récessif; les cellules des individus qui ont un trait anémique contiennent deux types d'hémoglobines, tandis que les cellules des

individus sains et des individus anémiques ne contiennent qu'un seul type d'hémoglobine. Chaque type de protéine pourrait donc être traduit à partir d'allèles différents du même gène. Les recherches menées en génétique classique et en biologie moléculaire, même si elles sont menées de manière indépendante, se confirment.

Une objection pourrait être faite à la défense d'un concept de maladie génétique qui s'appuierait sur cette double dimension de transmission héréditaire et d'explication de niveau moléculaire de certaines pathologies. Qu'en est-il en effet aujourd'hui des maladies *génétiques* complexes, dont l'hérédité est mal connue, et qui sont seulement associée à des facteurs de risques spécifiques ?

3 ⇨ Les statistiques sont aveugles

L'analyse de Magnus a également le défaut de ne pas prendre en compte l'articulation entre les études statistiques portant sur les facteurs de risque génétique et les recherches moléculaires, qui, on l'a vu, sont liée à la physiopathologie. On peut le comprendre: les méthodes statistiques permettent d'identifier des gènes-P dont les activités spécifiques, au niveau moléculaire, restent inconnue. Dans le cas des maladies qui se transmettent selon les lois de Mendel, le «gène qui cause la maladie» a donc souvent eu pour commencer une valeur prédictive et instrumentale. Aujourd'hui, l'analyse systématique du génome à la recherche de corrélations entre fréquence accrue de pathologie et marqueurs SNP, ou *Genome Wide Association Study* (GWAS), est une étude d'association: on cherche à mettre en évidence une différence de fréquence allélique au niveau d'un même locus chez des individus atteints ou non de maladies communes. Comme pour le gène-P, rien n'est connu encore une fois du mode d'action spécifique de cet allèle sur le développement de la pathologie. Magnus aurait-il donc raison de séparer les approches statistique et explicative ?

Avant de répondre à cette question, soulignons d'abord une première articulation entre études statistiques et études héréditaires (ce qui d'ailleurs est cohérent avec l'articulation des dimensions héréditaire et moléculaire des maladies génétiques). Pour étudier la composante génétiques des maladies multifactorielles à hérédité complexe, dont l'apparition dépend de l'interaction de plusieurs gènes et de facteurs environnementaux, il est nécessaire d'en démontrer la dimension héréditaire. Dans leur livre de référence, *Human Molecular Genetics*, Tom

STRACHAN et Andrew READ (2004) séparent l'étude des maladies génétiques en deux groupes. Un chapitre «Identifying human disease gene» est consacré aux méthodes qui permettent de localiser les gènes de maladies mendéliennes monogéniques. Un autre chapitre, «Mapping and identifying genes conferring susceptibility to complex diseases», traite des méthodes qui permettent d'identifier les gènes de susceptibilité et de prédisposition. Or, la manière dont les auteurs présentent la différence entre les maladies mendéliennes et les maladies complexes est très instructive. Ils expliquent en effet que personne ne contesterait l'idée selon laquelle une maladie est génétique si elle suit clairement un mode de transmission mendélien, qui suppose qu'un seul gène est transmis dans une famille. Par contre, pour les traits complexes, il est nécessaire *de faire la preuve que des facteurs génétiques sont impliqués dans le développement de la maladie*. Et pour cela, une recherche sur la transmission de la maladie dans les généalogies est nécessaire. Ce type de recherche a permis par exemple de distinguer deux formes différentes de la maladie d'Alzheimer : une forme familiale (FAD ou *Familial Alzheimer's Disease*) qui se transmet de manière autosomique dominante et une forme sporadique, liée à APOE e4. Trois gènes de prédisposition (PSEN1, PSEN2 et APP) confèrent un risque très élevé de développer la FAD précocement. La FAD est rare (5 %) ; la forme sporadique est la plus fréquente. Il existe alors plusieurs manières de travailler sur les maladies multifactorielles pour en calculer la composante génétique (pour une revue, cf. FEINGOLD 2005 et CAMPION 2001), mais trois étapes sont toujours nécessaires : «Montrer que la maladie est familiale, montrer que cette tendance familiale est due à des facteurs génétiques, et, enfin, identifier les gènes impliqués» (FEINGOLD 2005, p. 927). La recherche de déterminants génétiques repose donc sur les études du mode de transmission de la maladie dans les généalogies.

Finalement, l'ensemble des méthodes statistiques les plus efficaces prennent sens dans un contexte de recherche où l'hérédité est prise en compte. Les études d'agrégation familiale cherchent à savoir quelle est la prévalence de la maladie chez les apparentés par rapport à la prévalence dans la population. L'étude des jumeaux repose sur la comparaison de la concordance d'une maladie chez les jumeaux monozygotes et dizygotes. Les études de ségrégation consistent à collecter des arbres généalogiques et à modéliser le nombre de gènes impliqués. Pour les études d'association qui ne supposent pas, nous l'avons vu, d'étude d'ap-

parentés mais simplement de deux populations (atteint ou non atteint), une étude du sujet atteint et de ses deux parents complète l'analyse (FEINGOLD 2005, p. 930). Dans un article récent de *Nature* (MANOLIO *et al.* 2009), les auteurs expliquent que les GWAS ont permis d'identifier des centaines de variations génétiques associées à des maladies ou des traits humains complexes. Cependant, ces variants expliquent seulement une petite proportion de l'héritabilité : il est donc nécessaire de trouver des stratégies de recherche qui vont au-delà des GWAS pour en rendre compte.

La seconde articulation manquée par Magnus concerne les liens entre approches statistiques et explicatives des maladies. Les études d'association sont souvent accompagnées de recherches de gènes candidats, sélectionnés sur le critère de la cohérence entre leur fonction et les caractéristiques de la pathologie étudiée. On le comprend : les difficultés de l'étude des composantes génétiques des maladies complexes sont liées aux propriétés statistiques des études (par exemple, les problèmes du seuil de signification ou du choix des populations) et à la faiblesse du risque relatif conféré par chaque allèle de susceptibilité. Il faut donc essayer de trouver par ailleurs ce qui pourrait être modifié par ces allèles dans les voies biologiques caractéristiques des différentes pathologies étudiées.

Deux exemples permettent de mieux comprendre cette stratégie. Le premier est un article de Jean Dausset sur les systèmes d'histocompatibilité et la susceptibilité aux cancers (DAUSSET 1968). Ayant rapporté en 1958 les antigènes d'histocompatibilité au patrimoine héréditaire des individus, Dausset cherche à expliquer pourquoi certains cancers, comme les leucémies lymphoïdes chroniques, sont très fréquents dans certaines populations et presque absents dans d'autres. Lorsqu'il met en évidence, en 1968, une association entre des différences antigéniques caractérisant des populations de souris et leur résistance ou non à un virus leucémogène, la question qu'il pose immédiatement est la suivante : comment expliquer cette association ? Dans cette étude, la présence d'allèles spécifiques du système d'histocompatibilité H-2 est liée à la tendance des souris à résister au cancer ; mais un lien n'est pas encore une explication. Les « gènes de résistance » (DAUSSET 1968, p. 1397) situés au locus H-2 ou à côté de ce locus pourraient expliquer cette résistance d'une population de souris aux leucémies, mais aucune expérience ne « ne démontre directement l'intervention du locus H-2

lui-même» (*ibid.*, p. 1398). Enfin, d'après l'explication immunologique du phénomène de résistance, le concept de cause désignerait dans ce contexte une molécule qui empêcherait le virus de pénétrer dans la cellule. Mais la mise en évidence d'un rôle des produits des gènes de résistance dans les phénomènes de reconnaissance de l'antigène ou de pénétration du virus dans la cellule est, à cette époque, hors de portée des chercheurs. Cette expérience ne met donc en évidence qu'une corrélation entre certains allèles et une fréquence diminuée de contamination par le virus; elle ne démontre pas de causalité. Par contre, elle montre bien que les recherches statistiques prennent sens dans le contexte de tentatives d'explication qui concernent les effets des gènes sur les processus pathologiques.

Le second exemple est plus contemporain. Les difficultés rencontrées dans le cadre des études statistiques ont conduit les chercheurs à progresser dans la mise en évidence de relations plus adéquates entre des phénotypes mieux décrits et individués et des facteurs génétiques candidats au statut de cause. Dominique Campion rappelle que les analyses de liaison, qui reposent sur la connaissance du mode transmission d'une maladie, sont le plus souvent inadaptées pour l'étude des maladies complexes. Les études d'association, quant à elles, «trouvent l'essentiel de leur intérêt lorsqu'elles s'appliquent à des "gènes candidats"» (CAMPION 2001, p. 1139). Il affirme alors qu'un gène intervient toujours dans une voie biologique qu'il contribue à modifier. Selon lui, «pendant longtemps, il a été admis plus ou moins implicitement que maladies mendéliennes et multigéniques concernaient deux types de caractères biologiques radicalement différents» (*ibid.*, p. 1144). En réalité, dans les deux cas les gènes interviennent en modifiant des voies biologiques. Dans le cas des maladies monogéniques, la mutation d'un seul gène entraîne une perturbation majeure de la voie, qui a des conséquences fonctionnelles importantes. Dans le cas des maladies multigéniques, les modifications ne sont pas aussi sévères, mais les effets de plusieurs facteurs de risque impliqués dans une même voie biologique peuvent s'accumuler jusqu'à ce qu'un seuil soit dépassé, ce qui correspond à l'apparition de la pathologie. Quelle est donc la stratégie pour identifier de bons gènes candidats? C'est d'identifier, d'abord, les voies biologiques impliquées dans l'apparition d'une pathologie.

Cette analyse éclaire pourquoi le «réexamen» de la maladie de Huntington effectué par Magnus n'est pas si étonnant qu'il le prétend.

Cette maladie est liée à une répétition caractéristique de codons CAG dans une portion du chromosome 4. Or, on a montré qu'il est difficile de prédire si les individus qui ont entre 30 et 40 répétitions seront ou non malades. Cette difficulté est perturbante si l'on considère que «le gène de la maladie de Huntington» fonctionne à la manière du «tout ou rien». Mais si l'on admet que la maladie résulte de la perturbation d'une voie biologique et que les effets des mutations du gène sont quantitatifs, on comprend mieux pourquoi elle fonctionne par effets de seuil. L'expressivité variable des maladies monogéniques, comme la sévérité des symptômes ou leur moment d'apparition, est également mieux expliquée par la définition de seuils qui correspondraient aux interactions de la protéine avec les autres variables impliquées dans la voie biologique. On peut se référer à cette citation du grand généticien Victor McKusick : «L'identification moléculaire a révélé l'occurrence fréquente de "plusieurs à partir d'un" (des phénotypes multiples à partir de différentes mutations du même gène) et de "un à partir de plusieurs" (un même phénotype à partir de mutations dans deux ou plusieurs gènes). Les principes fondamentaux de la génétique clinique sont la pléiotropie, l'hétérogénéité génétique et la variation.» Pour la mucoviscidose, on recense ainsi plusieurs variants différents associées au développement de la maladie et dont les effets sont plus ou moins délétères (136 par exemple pour CFTR). La thalassémie est liée à des mutations de plusieurs gènes différents.

Pour définir les voies biologiques perturbées dans le contexte des maladies génétiques, la mise en évidence de sous-entités mendéliennes et l'utilisation de phénotypes intermédiaires sont des étapes souvent nécessaires. C'est le cas par exemple de maladies telles que la schizophrénie, où les difficultés à reproduire les études de liaison et d'association ont conduit les chercheurs à identifier des phénotypes intermédiaires associés à un déterminisme génétique plus lisible. Par exemple, CAMPION (2001) évoque les troubles des mouvements de poursuite oculaire lents et les troubles du filtrage sensoriel, deux endophénotypes associés à la schizophrénie. L'équipe de Maziade (MAZIADÉ *et al.* 2003) fait référence à des phénotypes «dimensionnels» (des agrégats de symptômes) ou des phénotypes neurocognitifs associés à la maladie et génétiquement moins complexes. Finalement, ce sont ici les études génétiques qui, par leur difficulté, orientent les chercheurs vers des nosographies différentes : la «schizophrénie» tend de plus en plus à être ramenée à des étiologies

différentes tandis que ses frontières avec les désordres bipolaires se troublent.

Pour autant, est-il justifié de considérer ces maladies multifactorielles comme des maladies «génétiques»? Il ne me semble pas qu'en se penchant sur la génétique des maladies complexes les chercheurs visent à démontrer que l'asthme ou les troubles bipolaires sont des maladies génétiques. Par contre, ils utilisent des stratégies de recherche qui leur permettent de mieux comprendre les facteurs génétiques impliqués dans ces maladies. Quel en est l'intérêt? Mieux comprendre le développement physiopathologique de la maladie et identifier des pistes thérapeutiques en fonction d'une meilleure compréhension des interactions entre gènes et environnement. Parfois, les études génétiques ont également un intérêt d'ordre diagnostique: le variant HLA-B27 sert ainsi à diagnostiquer la spondylarthrite ankylosante.

4 ⇨ Le choix de la génétique

Revenons une dernière fois à la question de savoir pourquoi, parmi toutes les causes possibles des maladies, les causes génétiques sont parfois privilégiées. Magnus termine son article sur l'idée suivante: les réponses épistémologiques à cette question sont insuffisantes. Reste donc à examiner l'usage du concept de maladie génétique. Il explique alors:

Étiqueter une maladie comme génétique revient à reconnaître implicitement que, pour cette maladie, la compréhension et les thérapies seront mieux servies par une recherche au niveau génétique. En d'autres termes, le cœur des classifications conceptuelles est un ensemble d'engagements concernant la meilleure façon d'allouer les ressources ainsi que la meilleure façon de pratiquer la science et la médecine. (MAGNUS 2004, p. 240.)

Or, selon Magnus, nous manquons cruellement d'arguments empiriques justifiant le financement de recherches génétiques. Par contre, cette remarque amène à la nécessité d'identifier les valeurs extérieures à la science qui conduiraient à privilégier ce type de recherches.

Pour préciser cette critique, j'aimerais reprendre l'analyse proposée par KITCHER (2000) sur les raisons qui amènent à privilégier, en biologie et en médecine, le déterminisme génétique, alors même que l'on reconnaît une influence de l'environnement sur le développement et les comportements des humains. Cet article est une réponse à Richard Lewontin qui a «fait le diagnostic des erreurs qui ont séduit des éru-

aits influents et leurs lecteurs à croire des slogans populaires sur les gènes et la destinée» (*ibid.*, p. 283). LEWONTIN (1992) critique la conception populaire du déterminisme génétique en montrant qu'elle ignore l'interaction entre les facteurs génétiques et non génétiques. C'est la raison pour laquelle l'extension de la catégorie de maladie génétique est socialement dangereuse : elle conduit à s'imaginer que ce qui est «génétique» est «inévitable» et, de proche en proche, conduit à considérer que la seule solution aux problèmes «génétiques» serait la sélection des individus sur des critères génétiques. Pour Lewontin, cette conception populaire a une source plus profonde : la biologie devrait, au moins en partie, être conceptualisée autrement. Dans la lignée de Lewontin, Susan Oyama par exemple défend la thèse selon laquelle les oppositions à partir desquelles les biologistes travaillent, par exemple les oppositions entre inné et acquis, ou entre gène et environnement, ne sont pas pertinentes (OYAMA 1985). Porter son attention sur le gène comme un facteur causal n'est qu'une abstraction de situations causales complexes qui donne injustement la priorité à certains déterminants du phénotype. Or, Kitcher entreprend de montrer que la conception interactionniste, qui sépare les gènes et l'environnement, ne doit pas pour autant être rejetée. Il affirme :

Le déterminisme génétique persiste non en raison d'une erreur subtile dans les idées conventionnelles concernant le caractère général de la causalité biologique, mais parce que les scientifiques qui étudient des traits compliqués dans des organismes complexes ont tendance à appliquer incorrectement des conceptions générales correctes. (KITCHER 2000, p. 284.)

Il défend alors une conception «démocratique» des causes environnementales et génétiques et explique pourquoi, en fait, les causes génétiques sont privilégiées par les scientifiques.

Pour démontrer sa thèse, Kitcher analyse un concept couramment utilisé en biologie, celui de norme de réaction. La stratégie qui soutend les recherches du déterminisme génétique de traits biologiques «commence par isoler certaines propriétés des organismes pour explorer leur impact causal, en considérant le phénotype comme le produit des contributions de séquences d'ADN d'un certain type d'un côté, et de *tout le reste* de l'autre. Elle continue en se demandant comment le phénotype varie quand les séquences d'ADN sont tenues constantes et quand les autres facteurs (la constitution cytoplasmique du zygote, les molécules

qui passent à travers la membrane plasmique, etc.) changent» (*ibid.*, p. 285). La norme de réaction du génotype est la représentation graphique de cette stratégie: le déterminisme génétique d'un phénotype est alors défini comme sa relative invariance, étant donné un seul génotype dans tous les environnements. Selon les opposants de la thèse interactionniste, les présupposés qui permettent de construire des normes de réaction, par exemple la distinction entre génotype et phénotype, doivent être repensés. Pourtant, Kitcher s'attache à montrer dans cet article que cet outil est scientifiquement valide. Il est légitime d'isoler certains facteurs causaux en les tenant constants pour chercher comment l'effet varie lorsque les autres facteurs sont modifiés.

Mais si cet outil est scientifiquement valide, il est cependant mal utilisé par les scientifiques. L'interactionnisme reconnaît en effet qu'il y a plusieurs causes impliquées dans le développement. Kitcher défend ainsi le principe de «démocratie causale», selon lequel il serait tout aussi justifié scientifiquement de proposer une analyse causale d'un facteur environnemental particulier en observant ce qui se passe lorsque le génotype est modifié: «Le principe de démocratie n'accorde aucun privilège particulier aux représentations qui mettent les gènes au premier plan» (KITCHER 2000, p. 290). Pourquoi, dès lors, les causes génétiques sont-elles, de fait, mises au premier plan par les biologistes?

La première raison invoquée par Kitcher est pragmatique: les chercheurs croient que les nouvelles technologies issues de la biologie moléculaire peuvent améliorer leur compréhension de certaines maladies et de certains comportements. Il explique ainsi que la génétique du comportement est porteuse de promesses, puisqu'il devrait être possible d'utiliser les techniques de séquençage de l'ADN pour identifier des allèles partagés par différents individus d'une population. Dans ce cas, si les facteurs environnementaux causalement pertinents étaient identifiés, ce qui constitue le plus grand problème de ce type d'étude, il serait possible d'étudier la variation des phénotypes en fonction d'un génotype tenu constant quand les facteurs environnementaux changent. Dans le domaine médical, il prend l'exemple des recherches sur l'alcoolisme et l'addiction, et explique les raisons pour lesquelles les scientifiques tentent d'identifier les causes génétiques:

Ils commencent avec les causes génétiques, non parce qu'ils sont convaincus qu'elles sont les plus importantes (que les normes de réaction de certains génotypes «d'addiction» seraient presque plates),

mais parce qu'ils veulent démêler la neurochimie et qu'ils considèrent que l'étude des génotypes est un fil directeur dans l'écheveau. (*Ibid.*, p. 295.)

Et c'est en un sens ce qui se passe, puisque ce type de stratégie permet d'identifier des entités nosologiques différentes pour une pathologie considérée auparavant comme une. Elle permet de mieux comprendre les modes de transmission de ces traits dans les familles et les éventuelles interactions géniques. La compréhension des pathologies est donc améliorée. Dans le cas évoqué par Kitcher, les chercheurs espèrent que la méthode de génétique inverse et la connaissance de la séquence d'ADN permettront de comprendre les changements moléculaires qui s'opèrent dans le cerveau. Ils pensent comprendre les détails moléculaires des interactions qui s'effectuent entre l'organisme et l'environnement et qui diffèrent entre les personnes dépendantes et les personnes non dépendantes. Ici, il n'est donc pas nécessaire de supposer une déformation sociale de la science qui s'appuierait sur des normes valorisées socialement pour justifier l'importance accordée aux recherches des déterminants génétiques des maladies.

Mais, selon Kitcher, on peut penser qu'une étude des normes de réaction des «allèles de la violence» serait encouragée socialement pour des raisons qui ne seraient pas scientifiques ou médicales. D'après le principe de démocratie causale, il serait possible de mener une étude où l'environnement serait tenu constant et où la variation des phénotypes serait observée en fonction de ce facteur causal. Dans ce cas, selon Kitcher, on pourrait voir que pour un même génotype, le phénotype varie selon les environnements et en conclure que l'environnement est un facteur causal du comportement violent. Selon Kitcher :

Dans une société qui tourne constamment et cyniquement le dos aux programmes qui pourraient aider les malchanceux et qui considère les impôts comme une forme de vol plutôt que comme des moyens nécessaires pour la coopération sociale, la recherche que j'ai ébauchée n'a pas d'intérêt évident. (*Ibid.*, p. 296.)

Au contraire, les études qui espèrent montrer qu'en gardant le génotype constant et en faisant varier les environnements, le phénotype ne varie pas (la norme de réaction est plate ou presque plate) ont plus de chance d'être financées. Dans ce cas, dépister ce génotype chez un individu permettrait alors d'affirmer qu'il a une prédisposition au comportement violent. Ce type d'études renforcerait donc l'idée que les solutions

sociales sont sans espoir: si un individu est prédisposé à la violence, il y a peu de chances que l'école par exemple lui permette d'échapper à ce déterminisme. Ce qui est génétique est considéré comme inévitable, alors même que tous les scientifiques s'accordent pour reconnaître qu'aucun trait n'est déterminé par les seuls gènes.

Cette manière de penser fait l'objet de nombreuses critiques depuis les années 1990. Le concept de «généticisation» est alors introduit comme une critique de l'extension d'une vision génétique de l'homme accompagnant les découvertes scientifiques. Abby LIPPMAN (1991) et Henk TEN HAVE (2001) désignent ainsi un processus qui touche la médecine et la société. Il consiste en une redéfinition des individus selon leurs gènes, le choix d'un nouveau langage pour décrire, interpréter et comprendre la vie humaine, et une explication des différences individuelles par des différences génétiques. La généticisation est déterministe et réductionniste. Quelles en sont les conséquences sur le concept de maladie génétique?

Premièrement, en médecine, l'extension de la catégorie de maladie génétique conduit à une perte de spécificité de l'adjectif «génétique». Mais si toutes les maladies sont considérées comme génétiques, alors le concept de maladie génétique ne peut plus avoir de sens nosographique. Ainsi, si la notion de terrain génétique peut avoir un sens, elle ne doit pas conduire à faire de toute maladie une «maladie génétique» au prétexte que des allèles sont associés à une plus grande susceptibilité à cette maladie.

Deuxièmement, la généticisation a des conséquences sur l'approche instrumentale des causes. En effet, dire de X qu'il est génétique, c'est, selon la critique de la généticisation, abandonner toute politique publique visant à améliorer les conditions environnementales, puisque «génétique» équivaut à «inévitable». En même temps, considérer une maladie comme génétique, c'est aussi penser que la prise en charge la plus facile se situerait au niveau génétique. Mais quelle est donc la prise en charge la plus commune pour les maladies génétiques? La sélection des naissances à partir de tests génétiques. On ne peut alors qu'être soucieux des conséquences possibles de ce processus. En effet, l'extension du concept de maladie génétique conduirait à une extension de pratiques sur lesquelles pèse un soupçon d'eugénisme. Dans les débats actuels, l'argument de la pente glissante est fréquemment invoqué pour souligner le danger de l'extension des pratiques de diagnostic prénatal

ou préimplantatoire dans le contexte de prédispositions génétiques. Par exemple, autoriser de tels diagnostics pour des maladies à révélation tardive et dont la probabilité d'apparition n'est pas égale à 1, comme les cancers du sein et des ovaires dits héréditaires ouvrirait selon certains la voie à la sélection des naissances sur des critères beaucoup plus triviaux. Il y a donc, de ce point de vue, un véritable enjeu à bien délimiter le concept de maladie génétique.

5 ➤ Conclusion

Trois critères fonctionnent donc ensemble pour définir une maladie comme génétique. Le premier concerne la causalité spécifique qui permet de comprendre le développement de la pathologie. Un gène est dans ce contexte une unité d'explication d'une modification spécifique d'une voie biologique importante dans le cadre d'une pathologie. Le second critère est le lien entre ces explications et la dimension héréditaire des maladies, approchée dans le cadre d'études statistiques. Le dernier critère est la perspective d'une prise en charge médicale préventive ou curative compatible avec des impératifs éthiques. L'articulation de ces critères pourrait alimenter alors de manière pertinente les réflexions sur le concept de maladie. Par exemple, l'homosexualité est retirée du DSM (*Diagnostic and Statistical Manual of Mental Disorders*) en 1973. Mais on peut poser la question suivante : si les pratiques homosexuelles avaient été corrélées à un ensemble de facteurs génétiques à cette époque, auraient-elles ainsi été retirées de la liste des désordres mentaux ?

➤ Références bibliographiques ◀

C

CAMPION D. (2001), «Dissection génétique des maladies à hérédité complexe», *Médecine/sciences*, 17 : 1139-1148.

D

DAUSSET J. (1968), «Les systèmes d'histocompatibilité et la susceptibilité au cancer», *La Presse médicale*, 76(28) : 1397-1400.

F

FEINGOLD J. (2005), «Maladies multifactorielles : un cauchemar pour le généticien», *Médecine/sciences*, 11(21) : 927-933.

FELDMAN D. & TAUBER A. (1997), "Sickle cell anemia: reexamining the first molecular disease", *Bulletin of the History of Medicine*, 71(4) : 623-650.

G

GANNETT L. (1999), "What's in a Cause? The Pragmatic Dimensions of Genetic Explanations", *Biology and Philosophy*, 14: 349-374.

GIFFORD F. (1990), "Genetic traits", *Biology and philosophy*, 5(3): 327-347.

H

HESSLOW G. (1984), "What is a genetic disease? On the relative importance of causes", in L. Nordenfelt & B.I.B. Lindahl (eds.), *Health, Disease and Causal Explanation in Medicine*, Reidel: 183-193.

HULL R. (1979), "Why Genetic Disease?", in Capron *et al.* (eds.), *Genetic Counseling: Facts, Values and Norms*, Alan R. Liss: 57-69.

I

INGRAM V.M. (1957), "Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin", *Nature*, 180: 326-328.

INGRAM V.M. & STRETTON A.O. (1959), "Genetic basis of the thalassemia diseases", *Nature*, 184: 1903-1909.

K

KITCHER P. (1994), "Who's afraid of the human Genome Project", in D.L. Hull & M. Ruse (eds.), *The Philosophy of biology*, Oxford University Press, 1998: 522-535.

KITCHER P. (2000), "Battling the undead. How (and how not) to resist genetic determinism", in P. Kitcher, *In Mendel's mirror. Philosophical reflections on biology*, Oxford University Press, 2003: 283-300.

L

LAMBERT G. (2009), *Vérole, cancer & Cie. La société des maladies*, Seuil.

LEWONTIN R. (1992), *Biology as Ideology. The doctrine of DNA*, Richard Harper Perennial.

LIPPMAN A. (1991), "Prenatal genetic testing and screening: constructing needs and reinforcing inequities", *American Journal of Law and Medicine*, 17: 15-50.

M

MAGNUS D. (2004), "The Concept of Genetic Disease", in A. Caplan, J. McCartney & D. Sisti (eds.), *Health, Disease, and Illness*, Georgetown University Press: 233-242.

MANOLIO T.A., COLLINS F.S., COX N.J., GOLDSTEIN D.B. & HINDORFF L.A. (2009), "Finding the missing heritability of complex diseases", *Nature*, 461: 747-753.

MAZIADÉ M., MERETTE M., CHAGNON Y.-C. & ROY M.-C. (2003), «Génétique de la schizophrénie et de la maladie bipolaire», *Médecine/sciences*, 19(10): 960-966.

MOSS L. (2004), *What genes can't do*, MIT Press.

N

NEEL J.V. (1949), "The inheritance of sickle cell anemia", *Science*, 110: 64-66.

O

OYAMA S. (1985), *The Ontogeny of information. Developmental system and evolution*, Cambridge University Press.

P

PAULING L., ITANO H.A., SINGER S.J. & WELLS I.C. (1949), "Sickle cell anemia, a molecular disease", *Science*, 110: 543-548.

R

RIORDAN J.R., ROMMENS J.M., KEREM B., ALON N., ROZMAHEL R. & GRZELCAK Z. (1989), "Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA", *Science*, 245: 1066-1073.

S

SOBER E. (2001), "The meaning of genetic causation", in A. Buchanan, D. Brock, N. Daniels & D. Wikler, *From chance to choice. Genetics and justice*, Cambridge University Press : 347-370.

STERELNY K. & KITCHER P. (1988), "The Return of the Gene", *The Journal of Philosophy*, 85: 339-361.

STRACHAN T. & READ A.P. (2004), *Human Molecular Genetics 3*, Garland Science.

T

TEN HAVE H.A.M.J. (2001), "Genetics and culture: the geneticization thesis", *Medicine, Health care and philosophy*, 4: 295-304.

TSUI L.-C., BUCHWALD M., BARKER D., BRAMAN J.C. & KNOWLTON R. (1985), "Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker", *Science*, 230: 1054-1057.

☉Notice biographique☉

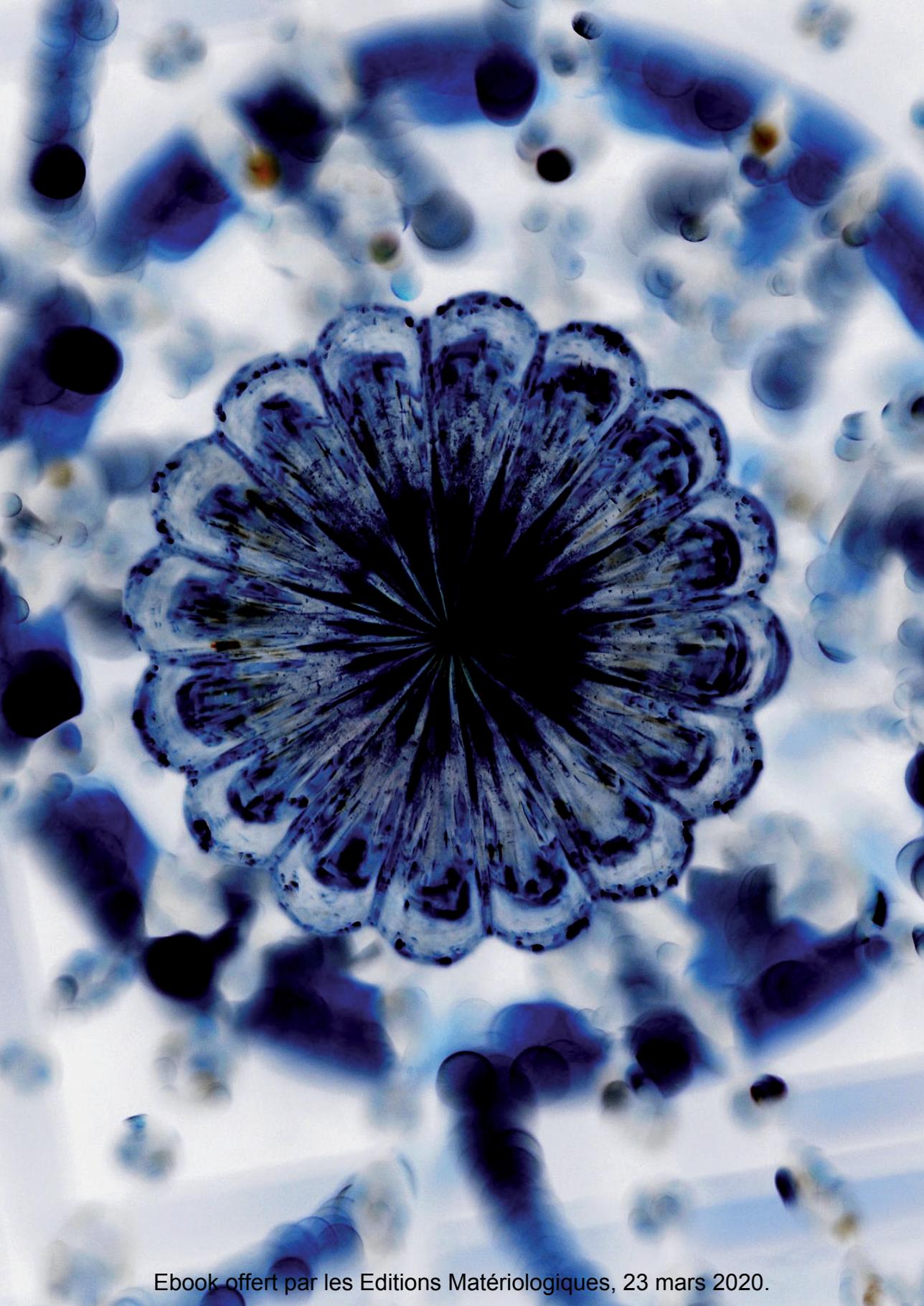
Catherine Dekeuwer est maître de conférences à la faculté de philosophie de l'Université Jean Moulin Lyon 3 et chercheur associé au Centre de recherche Sens, éthique et société (UMR 8137, Université Paris Descartes). Travaux en relation avec le thème développé dans « Définir la maladie génétique » : « Liberté de choix et destins individuels. Examen des concepts et des problèmes éthiques impliqués par la médecine prédictive », thèse de doctorat de l'Université Paris 1 Panthéon-Sorbonne, 2006 ; « Sur quelques raisons historiques de l'ambiguïté de prédire en médecine prédictive », in Ilario Rossi (dir.), *Prédire et prévoir la maladie. De la divination au pronostic*, Aux Lieux d'Être, 2007 ; « Imaginaire de la santé parfaite et idéologie génétique », *Diagonale-Phi*, n° 3, 2009 ; « Maladies d'Alzheimer, génétique et concepts d'autonomie », in N. Kopp, C. Thomas-Antérion, M.-P. Réthy, J.-P. Pierron, *Alzheimer : autonomie, pensée et volonté*, Belles lettres, 2010.

Avertissement. Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur et des auteurs. Les textes et illustrations figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage personnel, scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document selon ce modèle :

Catherine Dekeuwer, « Définir la maladie génétique », in Gérard Lambert et Marc Silberstein (dir.), Matière première. Revue d'épistémologie [en ligne]. Nouvelle série, N° 1/2010 : Épistémologie de la médecine et de la santé, Éditions Matériologiques. Mis en ligne le 7 octobre 2010.

URL : www.materiologiques.com

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.





James A. Marcum **Cancer : complexity, causation, and systems biology**¹

During the early part of the twenty-first century, especially after the completion of the Human Genome Project and the introduction of high throughput genomics and proteomics, biologists realized that biological entities and processes exhibit tremendous and often bewildering complexity in which causation is simply not linear at times but also multifaceted and reciprocal. In discussing cellular complexity, for example, Henri Atlan notes, “The great challenge today is to succeed in managing the enormous amount of experimental data on correlations of molecular structures and interactions with cellular functions” (ATLAN 2003, p. 2). That challenge not only involves the formulation of “organizational laws”, according to Atlan, but also the development of theoretical systems to interpret the experimental data and to guide additional experimental research in the investigation of biological complexity and causal pathways. Moreover, biologists must broaden their horizons beyond molecular and cellular complexity and causal pathways, he insists, to include tissues, organs, and organ systems, as well as the organism and its environment (both physical and social).

To meet Atlan’s challenge, some biologists argue that twenty-first century biologists must replace a constrained reductionism approach

1. Article initialement publié dans *Medicina & Storia*, vol. 2/2009. Nous remercions chaleureusement Bernardino Fantini, qui nous a autorisé à reproduire ici ce texte. (NDR.)

with an unrestricted antireductionism approach like organicism or holism (MESAROVIC *et al.* 2004, VAN DE VIJVER *et al.* 2003, VAN REGENMORTELM 2004). For example, Marc van Regenmortel asserts that reductionism “has reached its limit. Biological systems”, he adds, “are extremely complex and have emergent properties that cannot be explained, or even predicted, by studying their individual parts” (*ibid.*, p. 1016). The scientific investigation of cancer over the last several decades witnessed such a transition in its fundamental assumptions, with an antireductionism approach now challenging the traditional genetic reductionism approach (MALATERRE 2007, MARCUM 2005). Today, however, a systems biology approach to the disease is vying to eclipse the reductionism and antireductionism approaches through a synthesis of both (MARCUM 2009).

In this paper, I first examine the reductionism approach that underlies scientific research into the complexity and causal pathways of carcinogenesis, especially the work of Robert Weinberg and colleagues at the Whitehead Institute in Cambridge, Massachusetts. To illustrate this approach, I examine contemporary research on the genetics of breast carcinogenesis. I then discuss the challenge to this approach by those advocating an antireductionism approach, particularly the organicism approach of Carlos Sonnenschein and Ana Soto from Tufts Medical School in Boston. I next explore the recent systems biology approach to cancer complexity and causal pathways, as illustrated by the research of Mina Bissell and coworkers at the Ernest Orlando Lawrence Berkeley National Laboratory in California. I conclude by discussing the questions of whether systems biology represents a synthesis of the reductionism and antireductionism approaches for investigating the complexity and causal pathways in carcinogenesis and whether it represents a novel approach.

1 ↻ Reductionism and antireductionism

Reductionism served modern science admirably for at least the last several hundred years and became the predominant assumption underlying scientific research in recent times (DUPRÉ 2000, MARCUM & VERSCHUUREN 1986). Although some scientists often propose antireductionism as an alternative to surpass the perceived limits or constraints of a reductionism approach to complex phenomena, reductionism always prevails as the means to progress in terms of assuming how best to scru-

tinize the causal relations of and to explain complex natural phenomena. Recently, however, the immense complexity of these phenomena has put a strain on the appropriateness and sufficiency of the reductionism approach. In this section, I use the investigation and explanation of cancer in terms of mutated genes to explore the reductionism approach, followed by a recent challenge of an antireductionism approach.

1.1 ↻ Reductionism

Contemporary science relies upon several types of reductionism, including theoretical, ontological, and methodological reductionism, to study and explain natural phenomena (BOOGERD *et al.* 2007, DUPRÉ 1993, 2000, MARCUM & VERSCHUUREN 1986). Theoretical reductionism is the best-known and most prevalent type of reductionism. It simplifies theories by reducing terms from a higher-level theory, e.g. a biological theory, to terms from a lower-level theory, e.g. theories in chemistry or physics. It generally moves from non-physicochemical terms to physicochemical terms. Ontological reductionism simplifies phenomena, by reducing higher-order or complex phenomena, e.g. a biological organism, with respect to lower-order or simple components, e.g. molecules and atoms. In other words, it simplifies phenomena by accepting the possibility that new properties for higher-order or complex phenomena are reducible to the properties of the elements composing them. Finally, methodological reductionism simplifies research by reducing – generally through technological means – higher-order or complex entities with respect to their elements at a lower order. Researchers utilize this type of reductionism to investigate just the elements or parts and not the complex entity as a whole.

Reductionism involves bottom-up or upward causal chains of events, entities, or forces, often connected in a linear fashion (BRUGGEMAN *et al.* 2002, DENT 1999, DUPRÉ 1993, MESEAROVIS *et al.* 2004, NOBLE 2003, VAN REGENMORTEL 2004). According to proponents of this approach, elemental components and their properties are responsible causally for the appearance or existence of complex phenomena and their properties. In other words, higher-order phenomena and their properties are the result of lower-order phenomena and their properties and are accountable for in terms of them. For example, biological organisms and their properties are the result of or dictated by their separate parts and properties, e.g. inorganic and organic molecules, and their properties account for cells and their properties, which in turn account for tissues and their proper-

ties, and so on. Consequently, researchers need not assign properties to or invoke properties for these phenomena as a whole, since the properties of the whole does not exceed the collective or causal agency of the individual components that compose it. The elemental components and their properties are causally sufficient to account for or to generate higher-order or complex phenomena and their properties.

Contemporary cancer research illustrates the reductionism approach in terms of theoretical, ontological, methodological reductionism and with respect to investigating and explaining complex biological phenomena in terms of simpler components. Although scientists have proposed various theories to account for the cause of cancer, the most influential one is the reduction of cancer to a set of genes and their products, especially the oncogene gene theory of cancer (MORANGE 1993, WEINBERG 1988a). This theory is grounded in the somatic mutation theory (SMT) attributed originally to Theodor BOVERI (1914). It gained community wide acceptance in the 1980s, when Robert Weinberg and colleagues demonstrated that a single point mutation in RAS, an oncogene, was adequate to transform cells from a normal phenotype to a cancerous one (WEINBERG 1988a). Subsequent research demonstrated that mutation of other genes, such as the tumor suppressor gene, retinoblastoma, was also required for explaining carcinogenesis (WEINBERG 1988b). According to WEINBERG (1998), mutated genes are the material cause of cancer. The triumph of this approach resulted in the production of a cancer cell from just a handful of mutated genes (HAHN *et al.* 1999).

Breast cancer is an apposite example of a molecular disease, since contemporary biologists reduce its etiology to a limited number of causal components (SUTER & MARCUM 2007). These components comprise various molecular mechanisms, involving cell growth and proliferation, regulated by different genes. A select number of genes altered in terms of carcinogenesis include the oncogenes HER2, which is an epidermal growth factor receptor gene, c-MYC, which transcribes a transcription factor involved in the regulation of cellular proliferation, apoptosis (programmed cell death), differentiation, and metabolism, and HRAS, which transcribes a kinase involved in cell apoptosis, differentiation, motility, and proliferation. Other oncogenes include the estrogen receptor gene and the cell cyclin D1 and E genes, which are involved in cell cycle regulation. Mutation of tumor suppressor genes is also critical for breast carcinogenesis, including TP53, RB1, and PTEN, which function

as gatekeepers, and especially the breast cancer susceptibility genes BRCA1 and BRCA2, which function as caretakers².

In a 2000 millennial issue of the journal *Cell*, Weinberg, along with Douglas Hanahan, proposed an inclusive reductionism paradigm to guide twenty-first century cancer research (HANAHAN & WEINBERG 2000). They presented the idea of heterotypic cell biology, in which tumor cells recruit normal cells to form a mature tumor. Rather than explaining cancer with just a handful of mutated genes, which might be adequate for initiating a tissue along the path to tumorigenesis, Weinberg and Hanahan argued that cancer, as a mature disease, is complex and multifaceted and exhibits at least six “hallmarks”. These hallmarks include specifically self-sufficiency in growth signals, insensitivity to antigrowth or growth-inhibitory signals, limitless replicative potential, evading apoptosis, sustained angiogenesis (blood vessel generation), and tissue invasion and metastasis. The cancer community received this new paradigm readily, which enlarged the horizon of reductionism to include additional components in carcinogenesis. Although the paradigm still maintains the cause of cancer is genetic mutation within normal cells, such an enlarged reductionism allows researchers to ask even broader questions about the cause of the disease and to utilize a range of practices to solve its complex nature.

Weinberg and coworkers recently extended the idea of complex heterotypic interactions implicated in tumorigenesis to include cell-microenvironment exchanges. Specifically, they examined the epithelial-stromal interactions in breast cancer development, using a “cleared” murine mammary fat pad model (KUPERWASSER *et al.* 2004). The model consists of epithelial cells removed from the fat pad leaving only the connective tissue stroma. By introducing genes for either hepatocyte growth factor or type β 1 transforming growth factor into human fibroblasts, Weinberg and coworkers demonstrated that these genetically altered fibroblasts support the growth of premalignant or malignant epithelial cells from normal human breast tissue, while normal fibroblasts do not. The study, the authors contend, “underscores the critical role of heterotypic interac-

2. Gatekeeper genes stop the cell cycle until DNA damaged during cell division is repaired, while caretaker genes are involved in the repair of damaged DNA, e.g. damage induced through mutation.

tions in human breast development and carcinogenesis” (KUPERWASSER *et al.* 2004, p. 4966).

Daniel Medina, from the Baylor College of Medicine, comments on the above study accordingly: “In the foreseeable future, researchers will be able to develop transplantable cell lines of normal and premalignant human breast epithelia to test the effect of chemopreventive agents on growth and progression and to use these stable cell populations to examine the molecular and genetic basis for premalignant progression. Experiments that have been only dreams are now possible because of the innovative and pioneering experiments of this research team” (MEDINA 2004, p. 4724).

Weinberg and colleagues subsequently utilized their model to show that carcinoma-associated fibroblasts (CAFs) from invasive human breast cancer secrete stromal cell-derived factor 1 (SDF-1) (ORIMO *et al.* 2005). SDF-1 assists in recruiting endothelial cells thereby promoting tumor angiogenesis. Moreover, it promotes tumor growth *via* the α -chemokine CXCR4 receptor located on tumor cells. Notably, they assert that these “experiments do not address how CAFs acquire these unique phenotypes”, although they conclude that inhibition of these fibroblast phenotypes may provide an effective means for treating breast cancer (ORIMO *et al.* 2005, p. 345)³. Weinberg and his coworker, Tyler Jacks, realize that research on the role of the stroma or the extracellular matrix in carcinogenesis “begins to build the long-needed bridge between intracellular signaling biochemistry and the changes in morphology that are observed as cells advance progressively down the road toward malignancy” (JACKS & WEINBERG 2002, p. 925).

1.2 ➔ Antireductionism

Critics of the reductionism approach to carcinogenesis claim that cancer is too complex a disease to assume just genetic or simple molecular components as causal agents, especially in terms of bottom-up or upward causation (MOSS 2003, PROSTATE CANCER COMPLEXITY WORKING GROUP 2003, ROOT-BERNSTEIN 1999). Rather, these critics claim that higher-order entities and their organization or structure are responsible

3. Although Weinberg and colleagues are unclear on the exact mechanism of how CAFs acquire their phenotype, they still assume it involves genetic mutation.

for the origin of the disease and generally espouse an antireductionism approach, such as organicism or holism. Organicism, in particular, is an idea utilizing organic unity or a whole to explain biological processes at the level of higher-order entities and their properties rather than simply invoking their elemental composition (MARCUM & VERSCHUUREN 1986).

Importantly, organicists do not espouse vitalism, especially the variety that imagines an *élan vital* or an entelechy. Rather, they insist that the compilation of the parts as a whole is vital for examining and explaining complex phenomena, especially with respect to informational content or emergent entities and their properties (EL-HANI & EMMECHE 2000, FOSS 2002). Consequently, information and entities emerge that cannot be derived or deduced from investigating just the component parts in isolation; rather, only when scientists examine the whole can they explain emergent information or properties of complex phenomena. In this approach, as in many but not all antireductionism approaches, causation proceeds from top to bottom or downwards⁴. Consequently, carcinogenesis is the result not of lower-order perturbations, such as mutated genes, but of perturbations of higher-order entities and their properties, such as disrupted or disorganized tissues or organs.

Carlos Sonnenschein and Ana Soto are probably the best-known advocates of organicism in cancer research. SONNENSCHN & SOTO (1999) proposed a theory of carcinogenesis, which they call the tissue organization field theory (TOFT). According to TOFT, disruption in tissue and its structure or architecture – and not mutated genes – is the cause of cancer. For these researchers, mutated genes are the result, and not the cause, of the disease's disrupted normal tissue architecture. SOTO & SONNENSCHN (2004) assert that the SMT is a futile theory and cannot explain carcinogenesis, because it cannot account for a number of experimental anomalies and it consists of several questionable assumptions. For example, one of – if not the most critical – assumptions of SMT,

4. Claus Emmeche and colleagues demarcate three types of downward causation (EMMECHE *et al.* 2001). Weak downward causation refers to an organizing principle in which a higher-order phenomenon is composed of lower-order phenomena but not reducible to them. Medium causation involves functional causation in which a higher-order phenomenon comprises an emergent phenomenon. Strong downward causation is efficient in nature such that a higher-order phenomenon embodies an ontologically independent phenomenon.

according to Soto and Sonnenschein, is that quiescence is the default state for the cell cycle. In terms of TOFT, based on their experimental results, they assert that proliferation is the default state. Since the default state of the cell cycle is proliferation, then modification of the tissue and its microenvironment is adequate, they allege, for cells to continue to proliferate uncontrollably during carcinogenesis.

Sonnenschein, Soto, and collaborators supply experimental evidence to justify TOFT, especially that a modified stroma is crucial for mammary gland tumorigenesis. Specifically, they investigated the role of the carcinogen, N-nitroso-methyl urea, using a “cleared” rat mammary fat pad model (MAFFINI *et al.* 2004). They found that the carcinogen facilitates carcinogenesis through modification of the mammary epithelial cell stroma and not presumably through mutation of the Ha-RAS-1 oncogene. From these results, they inferred that the stroma – not the oncogene – is the “target” for the carcinogen in breast carcinogenesis. “From the TOFT perspective”, note Soto and Sonnenschein in later commentary, “the effects of carcinogens on subcellular structures and organelles (including genomic mutations), while variably deleterious to each cell in the host, are not viewed as directly responsible for the development of neoplasms” (SOTO & SONNENSCHNEIN 2006, p. 372). Valerie WEAVER & Penney GILBERT (2004), in remarks on the above study, concede that the evidence appears to support TOFT; however, they demand further analysis of the observed tumors to determine their “metastatic potential”⁵.

In an attempt to include the literature in defense of TOFT, SONNENSCHNEIN & SOTO (2005, 2006) cite evidence from several studies. Scott Cramer and colleagues, from Wake Forest University School of Medicine, provide the first study (BARCLAY *et al.* 2005). They investigated the role of stromal cells in the induction of prostate cancer by taking stromal cells from normal, benign, and cancerous prostate tissue and recombining them with a human prostatic epithelial cell line. They then grafted these recombinations separately into the renal capsules of nude

5. Another troubling issue is that Sonnenschein, Soto, and collaborators did not show whether the mutagen modifies the extracellular matrix directly or whether it mutates fibroblast genes. In addition, MEDINA & KITRELL (2005) in a similar study report that the target of the carcinogen, dimethylbenzanthracene, is not the stroma but rather mammary epithelial cell genes.

mice. Normal stromal cells and the epithelial cell line did not produce grafts, upon histological analysis of the xenografts, while stromal cells from benign tissue and the epithelial cell line produced organized but small grafts. Stromal cells from cancerous tissue and the epithelial cell line, however, produced highly disorganized grafts, which swarmed the adjacent renal tissue. From these results, the investigators claimed that these data “support a growing body of work from a number of different epithelial tumor systems demonstrating that the tumor microenvironment and specially, epithelial-mesenchymal interactions are critically important for tumorigenesis” (BARCLAY *et al.* 2005, p. 17).

Pepper Schedin and coworkers, from University of Colorado Health Science Center, provide Sonnenschein and Soto with their second study to support TOFT (McDANIEL *et al.* 2006). These researchers examined the role of postlactational involution in breast cancer by isolating the extracellular matrix from nulliparous and postlactational rats and combining it with a normal mammary epithelial cell line. Cell culture analysis of these recombinants demonstrated that the nulliparous matrix supports ductal organization of epithelial cells while it inhibits invasion of a tumor cell line. Further, postlactational matrix does not support ductal organization of epithelial cells and does not inhibit invasion of a tumor cell line. Subsequent experiments in which the researchers combined the tumor cell line with different matrices and injected these combinations into the mammary fat pads of nude mice showed an increased metastasis for the postlactational matrix and tumor cell line combination to kidney, liver, and lung tissue, contrasted to the nulliparous matrix and tumor cell line combination. These investigators propose that the above results suggest a role for the postlactational matrix in tumorigenesis and metastasis. However, Sonnenschein and Soto make a startling statement concerning the significance of these results, concluding that the results of Schedin and coworkers, along with the results of others, “indicate that an explanation of carcinogenesis and metastasis at the cellular or subcellular [i.e. genetic] levels is incomplete, if not plain irrelevant” (SONNENSCHNEIN & SOTO 2006, p. 364).

2 ⇨ Systems biology

Weaver and Gilbert, remarking on Sonnenschein and Soto’s work (MAFFINI *et al.* 2004), declare, “from a practical perspective the tissue organization field theory of carcinogenesis might just converge or collide

with the somatic mutation theory” (WEAVER & GILBERT 2004, p. 1290). The two theories certainly collide with one another in that they each identify a different level of organization for the causal origin of cancer, and accept dissimilar core assumptions about the proliferative state of the cell cycle. Advocates of the modified SMT locate the cause of cancer at the gene level and assume the default state of the cell cycle is quiescence. According to SONNENSCHN & SOTO (1999) and their TOFT, causation of cancer is at the tissue level and the default state of the cell cycle is proliferation. However, another approach to carcinogenesis is gaining prominence, within the past decade, in the cancer community – systems biology – in which the two theories appear to be converging (HORNBERG *et al.* 2005, KHALIL & HILL 2004, LIU *et al.* 2006). The systems biology approach avoids locating causation within a single, hierarchical level, so that the flow of causation is in one direction only – whether bottom-up or top-down (BRUGGEMAN & WESTERHOFF 2006). Rather, that approach advocates a flow of causation that is bidirectionally or reciprocal in nature.

Fabio Grizzi and colleagues, for instance, promote a kind of causation for the systems biology approach that they call “multiscale causality” (GRIZZI & CHIRIVA-INTERNATI 2006, GRIZZI *et al.* 2006). According to this kind of causality, “the observed phenomenon at each scale has structural and behavioral properties that do not exist at lower or higher organizational levels” (GRIZZI & CHIRIVA-INTERNATI 2006, p. 8). As a result, rather than situating causality at one particular level within a hierarchy and the causal chain flowing from that level to another, multiscale causality associates causation at different levels concomitantly. Causation, then, may flow in various directions simultaneously and may even be nonlinear⁶. Consequently, the flow of causation is bidirectional, at the least, and potentially reciprocal⁷. Such causation is possible because

-
6. Nonlinear causation consists of a causal mechanism triggering discontinuous change, especially change observed in dynamical nonequilibrium or chaotic systems (SCOTT 2007). Some biologists propose this kind of causation as a significant component in disease causation, especially cancer (JANECKA 2007, KARTHIKEYAN 2005).
 7. Bidirectional causation does not necessitate intersection of causal pathways, while reciprocal causation does. For bidirectional causation, the causation moves in one direction without eliciting a causal response from the affected level, while in reciprocal causation it does.

new phenomena and their properties emerge, which cannot be reduced to lower-order phenomena. These novel higher-order phenomena and properties demonstrate causal efficacy, with respect to lower-order phenomena⁸. Finally, this kind of causation is a property of a system and, thereby, its position is not precisely determinable at any specific hierarchical level.

The approach of Mina Bissell and colleagues illustrates well the systems biology approach to carcinogenesis⁹. They have proposed a dynamic reciprocity theory (DCT) to explain how the configuration of the extracellular matrix helps to guide gene expression and thereby to control the normal structure and function of tissues, especially the development of mammary glands (BISSELL & AGGELER 1987, NELSON & BISSELL 2005). Bissell and colleagues use DCT to address the question: “How do tissues and organs maintain homeostasis, and how do cells within tissues lose or overcome these controls in cancer?” (BISSELL *et al.* 2003, p. 753). According to DCT, tissue homeostasis relies on reciprocal signaling between the different levels of organization, especially between the extracellular matrix and nucleus (SPENCER *et al.*, 2007). Accordingly, carcinogenesis is an interruption in that signaling: “A dynamic and reciprocal relationship between the genetically damaged cells in a tissue and their microenvironment, whereby damaged cells modify their environment, which, in turn, in a vicious cycle, brings about more pathological behavior from the cells” (BISSELL & LABARGE 2005, p. 21).

Originally, Bissell and colleagues considered the mammary epithelial cell and its extracellular matrix as the functional unit of the mammary gland. However, succeeding modeling based on additional experimental results of both normal development of and tumor formation within mammary glands redirected them to propose that the basic unit of function is the mammary acinus (BISSELL *et al.* 2003, NELSON & BISSELL 2005). The architecture of the acinus consists of luminal epithelial cells surrounded by a basement membrane and myoepithelial cells. In addition, the mature acinus also contains a connective tissue

8. For example, higher-order phenomena can constrain the function of lower-order phenomena.

9. Although systems biology has an extensive history, it unfortunately lacks a consensus definition and no current agreement concerning its nature is forthcoming (KIRSCHNER 2005, MARCUM 2009, TREWAVAS 2006).

stroma composed of collagen and elastin, immune cells, adipose cells, and fibroblasts, as well as blood vessels. Its structural organization or architecture is critical for the mammary gland's apposite functioning, which is compromised during the formation of breast cancer.

To investigate the role of the extracellular matrix and tissue architecture on mammary histogenesis and carcinogenesis, Bissell and collaborators developed a three-dimensional (3-D) culture system (SCHMEICHEL & BISSELL 2003, LEE *et al.* 2007). By growing cells in an extracellular matrix like Matrigel or one of the many components that make up the matrix like collagen, they achieved a 3-D structure. Rather than plating the cells on the surface of the culture vessel, the researchers embed them in the matrix. These conditions faithfully replicate *in vivo* conditions. Importantly, this culture system affords a method for augmenting mammary gland function compared to the traditional two-dimensional (2-D) cultural system. For example, compared to the 2-D system mammary epithelial cells synthesize and secrete 3-10 times more casein or milk protein under 3-D conditions.

Using the 3-D culture system, Bissell and coworkers demonstrated that in contrast to normal breast cells malignant cells do not form acini (WANG *et al.* 1998). Interestingly, utilization of inhibitors to specific cell adhesion molecules and growth factors, under 3-D but not 2-D *in vitro* conditions, reinstated the normal architecture, even in the presence of the malignant cells. Moreover, treatment of the malignant cells under 3-D conditions with cell adhesion antibodies resulted in the down regulation of the expression of the growth factor genes and *visa versa*. Comparable treatment of malignant cells under 2-D conditions, however, did not result in down regulation of these genes. Based on these experimental results, Bissell and colleagues proposed the following organizational principle: coupling growth factor and mammary epithelial cell adhesion pathways entails normal tissue architecture (BISSELL *et al.* 2002)¹⁰.

10. Menno HULSWIT (2006) claims that most examples of downward causation are not causal in terms of "bringing about" but rather in terms of explanation and determination. However, Bissell's organizational principle truly represents downward causation, since high-order structure is responsible for "bringing about" lower-order phenomena, i.e. tissue structure is causally determinate for coupling the two cellular pathways.

As Celeste Nelson and Bissell acknowledge, “we will not fully understand how the mammary gland maintains tissue-specificity until we develop culture models that fully recapitulate its full structural and functional complexity” (NELSON & BISSELL 2005, p. 6). To accomplish that aim, they propose introducing the fourth dimension of time to their experimental system. Specifically, they intend to use “smart polymers” and confocal microscopy to chart in real time normal histogenesis and carcinogenesis in engineered breast tissue. Bissell and collaborators also postulate an alternative approach to understanding the complexity of breast tissue histogenesis and carcinogenesis through use of an analogy of models to the World Wide Web and the Internet: “Exploiting parallels between biological tissues and such systems may yield useful tools for modeling the human breast acinus” (BISSELL *et al.* 2003, p. 759).

3 ⇨ Synthesis or paradigm shift ?

An important philosophical question arises as to whether the systems biology approach represents simply a synthesis or combination of the reductionism and antireductionism approaches or whether it represents a completely new approach or a paradigm shift for investigating and explaining the complexity of carcinogenesis, especially in terms of causation. Although Soto and Sonnenschein claim that the two theories of carcinogenesis are incompatible or even incommensurable, they acknowledge, “this does not mean that the data gathered from experiments based on the SMT cannot be interpreted in the light and context of the TOFT” (SOTO & SONNENSCHHEIN 2004, p. 1104). Even though they concede that alterations or mutations in genes can indirectly modify the extracellular matrix, still Soto and Sonnenschein insist that the modified environment is directly or eventually accountable for carcinogenesis: “The TOFT removes the gene from the driver’s seat (genetic determinism), and brings the organism and its ability to self-organize as the conceptual focus (organicism)” (*ibid.* 2004, p. 1105)¹¹.

11. SOTO & SONNENSCHHEIN (2004) do not clarify the confusion introduced here in terms of indirect modification of the extracellular matrix via mutated genes. It may be that what they mean is that mutated genes may modify the extracellular matrix, which then allows cells to express their default state of proliferation but these genes do not directly cause the cells to proliferate.

Commenting on Soto and Sonnenschein's TOFT, James Coffman argues that their organicistic theory "undervalues the importance of genes in development (in this case the development of cancer)" (COFFMAN 2005, p. 459). Although he proposes "a paradigm that is a synthesis between genetic determinism's thesis and organicism's antithesis", Coffman surpasses a simple addition or amalgamation of these two approaches and formulates a theory of carcinogenesis founded neither on reductionism nor on organicism but rather on integrated hierarchical levels (*ibid.*, p. 459). "Either extreme is an oversimplification that ultimately fails", insists Coffman, "as causal explanation" (*ibid.*). He cites the gene regulatory network of Eric Davidson and colleagues to account for sea urchin development, as an exemplary illustration of such integration between genes at a lower order and cell/tissue/organism at a higher order, which could account causally for tumorigenesis¹².

Soto and Sonnenschein, in countering Coffman's critique of their TOFT, maintain, "we advocate a hierarchical view of biology that recognizes the existence of emergent phenomena and their causative powers. In this view both top-down and bottom-up approaches are used" (SOTO & SONNENSCHNEIN 2005a, p. 460). They claim that an essential question in carcinogenesis is determining what causes what. They insist that the causal progression involved in carcinogenesis, given the historical dimension of normal tissue development, is circular or reciprocal in nature. In other words, as normal histogenesis unfolds cells and extracellular matrix interact in a reciprocal fashion, as in the examples of limb and tooth development, which they cite to support TOFT. Soto and Sonnenschein also uphold that "in biology several levels of explanation are necessary. Development and cancer will not be reduced to complex series of protein interactions, but rather a multivariant explanation will be required" (SOTO & SONNENSCHNEIN 2005b, p. 116).

However, SOTO & SONNENSCHNEIN (2005b) are not entirely comfortable with the idea that a genetic program is responsible for determining either organismal development or carcinogenesis. They find solace in Lenny Moss (2003), who claims that the rise of the genetic program is

12. COFFMAN (2006) proposes a theory of "developmental ascendancy" in which bottom-up causation is responsible for the development of immature tissues. Once developing tissues mature, however, emerging properties exert top-down control on further tissue development and differentiation.

an outcome of confounding the molecular gene with the Mendelian gene. For these researchers, the genetic program is simply another reserve, like nutrients or the microenvironment, for explaining organismal development or carcinogenesis. As Soto and Sonnenschein conclude, “the cancer phenotype is an adaptive, emergent phenomenon occurring at the tissue level of organization and is susceptible to being normalized” (SOTO & SONNENSCHHEIN 2005b, p. 113). Consequently, genes function in carcinogenesis but that function is constrained by the tissue’s (dys) function. Significantly, gene mutation is not responsible for cancer but rather disruption of the tissue environment of cells is. In a shocking statement to advocates of the modified SMT, Soto and Sonnenschein claim that mutated genes are simply “irrelevant to carcinogenesis” (SOTO & SONNENSCHHEIN 2006, p. 372).

Soto and Sonnenschein admit that a synthesis between the SMT and the TOFT is conceivable, but only after “a rigorous weeding-out process [in which] contradictions are resolved and both spurious ‘facts’ and wrong premises are recognized and dismissed” (SOTO & SONNENSCHHEIN 2005b, p. 114). They caution against a “premature synthesis that may lead to an ‘anything goes’ attitude where if results do not fit one hypothesis, they may fit its opposite or an *ad hoc* alternative one; in other words, nothing is rejected and everything is explained” (*ibid.*). Although Soto and Sonnenschein admit that a reciprocal model with respect to causation is possible, they maintain that the tissue and its architecture are the primary causative agent for carcinogenesis.

Others also advocate a synthesis, but one that transcends a simple union of reductionism and organicism. This synthesis results in a novel position that integrates both reductionism (as thesis) and organicism (as antithesis). This position is truly innovative and assigns no pride of place to any one hierarchical level. For example, Michael SPORN (2006) proposes a synthesis between what he calls the “dichotomies” of reductionism and holism. “We need to accept”, argues Sporn, “the limits of reductionism, which can be intensely useful and practical for analysis of short-term changes in linear systems that are not subject to rapid environmental perturbation, and realize the need to develop new circular, systems-oriented approaches to more complex problems” (*ibid.* p. 370). He gives an illustration of oncogenes, which do not always function in terms of turning on the cell cycle; rather, these genes can also function like tumor suppressor genes by turning off the cell cycle.

Importantly, Sporn's synthetic position takes into account the context in which phenomena transpire.

Bissell and LaBarge advocate a definition of cancer that is representative of a synthesis or integration of the different levels involved in carcinogenesis rather than just situating causation primarily to one level: "Cancer [is] a disease that must *simultaneously* subvert the microenvironmental controls as well as the genetic program" (BISSELL & LABARGE 2005, p. 17, emphasis added). Accordingly, cancer is not simply a disease limited to either genes or tissues but rather it is a disease of the system or organism (JANECKA 2007). Systems biology represents a novel approach to cancer, because different levels of organization are integrated to account for carcinogenesis rather than locating a causal agent within one specific level. Causation, for the systems biologist, is holistic and not simply linear or a modified linearity, such as feedback loops¹³. For example, carcinogenesis is the outcome of nonlinear interaction at several hierarchical levels in which each level interacts with another within a particular timeframe or given instance. Synergy is the result of this nonlinearity, with respect to causation.

Synergy is an important concept in systems biology and buttresses the contention that systems biology is more than simply the combination of the reductionistic and antireductionistic approaches. The concept refers to the interaction of at least two phenomena such that the combinatory outcome is greater than the sum of the individual outcomes of the phenomena (ROTHMAN 1974). In other words, the outcome is not the arithmetic sum but rather the exponential product of interacting phenomena. For example, Joshua Liao and co-workers demonstrated that genetically modified mammary epithelial cells, over-expressing both cyclin D1 and c-MYC, develop more aggressive and invasive cancer in SCID mice than genetically modified mammary cells, over-expressing either cyclin D1 or c-MYC separately (WANG *et al.* 2007). The concept of synergy also has significant implications for cancer therapy, in which clinicians treat patients with cancer drugs in combination rather than independently or sequentially (KENNY *et al.* 2007, SUTER & MARCUM 2007). For example, combinations of temsirolimus, a rapamycin

13. Holistic causation is the simultaneous causal relationship between wholes and parts, rather than the temporal sequence of causal events (PLACEK 2004).

(mTOR) inhibitor, and ERA-923, an estrogen receptor- α antagonist, demonstrate synergistic inhibition of tumorigenic mammary epithelial cell growth *in vitro* (SADLER *et al.* 2006).

4 ➤ Conclusion

Systems biology offers a novel means by which to revolutionize the biological sciences. The challenge in the twenty-first century for cancer researchers is to develop new metaphysical categories, especially with respect to causation, and to identify organizational rules, which complement or even possibly subsume the reductionist or genetic rules. In terms of methodological challenges, these researchers need to develop methodological techniques, such as 4-D and artificial tissue culture conditions and mathematical or computational and *in silico* models, to motivate and direct the cognitive and material elements of scientific practices that will capture more wholly the intricacy of carcinogenesis. However, to be truly revolutionary, cancer systems biologists must expand their horizon to include the complete system. The “big picture”, as Sporn christens it, is critical for investigating and elucidating the nature of a complex disease like cancer and especially for treating it. Importantly, this picture must also incorporate not only the biological level but also the social level, such as a patient’s lifestyle and society’s efforts to support the patient. “We need a new, holistic approach to prevention”, enthuses Sporn, “that will include both chemoprevention *and* a better life-style, with cooperative efforts from people in both areas, directed toward a better quality of life for the patient” (SPORN 2006, p. 371).

➤ Acknowledgements ◀

I thank especially Silvia Camporesi, Bernardino Fantini, Gérard Lambert, Michel Morange, and other members of the conference on Causation in Biomedical Sciences II Philosophical and Historical Aspects, held at Hôtel de Sonloup, Montreux, on 31 March-2 April, 2008, for insightful comments on an earlier draft of this paper. Baylor University supported my research through a summer sabbatical award and a University Research Committee grant.

Bibliography

A

ATLAN Henri (2003), 'The living cell as a paradigm for complex natural systems', *ComplexUs*, 1: 1-3.

B

BARCLAY Wendy W., WOODRUFF Ralph D., HALL M. Craig & CRAMER Scott D. (2005), 'A system for studying epithelial-stromal interactions reveals distinct inductive abilities of stromal cells from benign prostatic hyperplasia and prostate cancer', *Endocrinology*, 146: 13-18.

BISSELL Mina J. & AGGELER Judith (1987), 'Dynamic reciprocity: how do extracellular matrix and hormones direct gene expression?', *Progress in Clinical and Biological Research*, 249: 251-262.

BISSELL Mina J. & LABARGE Mark A. (2005), 'Context, tissue plasticity, and cancer: are tumor stem cells also regulated by the microenvironment?', *Cancer Cell*, 7: 17-23.

BISSELL Mina J., RADISKY Derek C., RIZKI Aylin, WEAVER Valerie M. & PETERSEN Ole W. (2002), 'The organizing principle: microenvironmental influences in the normal and malignant breast', *Differentiation*, 70: 537-546.

BISSELL Mina J., RIZKI Aylin & MIAN I. Saira (2003), 'Tissue architecture: the ultimate regulator of breast epithelial function', *Current Opinion in Cell Biology*, 15: 753-762.

BOOGERD Fred C., BRUGGEMAN Frank J., HOFMEYR Jan-Hendrik S. & WESTERHOFF Hans V. (2007), 'Towards philosophical foundations of systems biology: introduction', in *Systems biology: philosophical foundations*, Amsterdam, Elsevier: 3-19.

BOVERI Theodor (1914), *Zur Frage der Entstehung Maligner Tumoren*, Jena, Fischer.

BRUGGEMAN Frank J., & WESTERHOFF Hans V. (2006), 'The nature of systems biology', *Trends in Microbiology*, 15: 45-50.

BRUGGEMAN Frank J., WESTERHOFF Hans V. & BOOGERD Fred C. (2002), 'BioComplexity: a pluralist research strategy is necessary for a mechanistic explanation of the "live" state', *Philosophical Psychology*, 15: 411-440.

C

COFFMAN James A. (2005), 'On reductionism, organicism, somatic mutations and cancer', *BioEssays*, 27: 459.

COFFMAN James A. (2006), 'Developmental ascendancy: from bottom-up to top-down control', *Biological Theory*, 1: 165-178.

D

DENT Eric B. (1999), 'Complexity science: a worldview shift', *Emergence*, 1: 5-19.

DUPRÉ John (1993), *The disorder of things: metaphysical foundations of the disunity of science*, Cambridge, Harvard University Press.

DUPRÉ John (2000), 'Reductionism', in *A companion to the philosophy of science*, Oxford, Blackwell: 402-404.

E

EL-HANI CHARBEL N. & EMMECHE Claus (2000), 'On some theoretical grounds for an organism-centered biology: property emergence, supervenience, and downward causation', *Theory in Biosciences*, 119: 234-275.

EMMECHE Claus, KÖPPE Simo & STJERNFELT Frederick (2001), 'Levels, emergence, and three versions of downward causation', in *Downward causation: minds, bodies and matter*, Århus, Århus University Press: 13-34.

F

FOSS Laurence (2002), *The end of modern medicine: biomedical medicine under a microscope*, Albany, SUNY Press.

G

GRIZZI Fabio, DI IEVA Antonio, RUSSO Carlo, FREZZA Eldo E., COBOS Everardo, MUZZIO Pier C. & CHIRIVA-INTERNATI Maurizio (2006), 'Cancer initiation and progression: an unsimplified complexity', *Theoretical Biology and Medical Modeling*, 3: 37-42.

GRIZZI Fabio & CHIRIVA-INTERNATI Maurizio (2006), 'Cancer: looking for simplicity and finding complexity', *Cancer Cell International*, 6: 4-10.

H

HANAHAH Douglas & WEINBERG Robert A., 2000, 'The hallmarks of cancer', *Cell*, 100: 57-70.

HAHN William C., COUNTER Christopher M., LUNDBERG Ante S., BEIJERSBERGEN Roderick L., BROOKS Mary W. & WEINBERG Robert A. (1999), 'Creation of human tumour cells with defined genetic elements', *Nature*, 400: 464-468.

HORNBERG Jorrit J., BRUGGEMAN Frank J., WESTERHOFF Hans V. & LANKELMA Jan (2005), 'Cancer: a systems biology disease', *BioSystems*, 83: 81-90.

HULSWIT Menno (2006), 'How causal is downward causation?', *Journal of General Philosophy of Science*, 36: 261-287.

J

JACKS Tyler & WEINBERG Robert A. (2002), 'Taking the study of cancer cell survival to new dimension', *Cell*, 111: 923-925.

JANECKA IVO P. (2007), 'Cancer control through principles of systems science, complexity, and chaos theory: a model', *International Journal of Medical Sciences*, 4: 164-173.

K

KARTHIKEYAN Ganesan (2005), 'Is disease causation random?', *Lancet*, 366: 25-26.

KENNY Paraic A., LEE Genee Y. & BISSELL Mina J. (2007), 'Targeting tumor microenvironment', *Frontiers in Bioscience*, 12: 3468-3474.

KHALIL Iya G. & HILL C. (2005), 'Systems biology for cancer', *Current Opinion in Oncology*, 17: 44-48.

KIRSCHNER Marc W. (2005), 'The meaning of systems biology', *Cell*, 121: 503-504.

KUPERWASSER Charlotte, CHAVARRIA Tony, WU Min, MAGRANE Greg, GRAY Joe W., CAREY Loucinda, RICHARDSON Andrea & WEINBERG Robert A. (2004), 'Reconstruction of functionally normal and malignant human breast tissues in mice', *Proceedings of the National Academy of Science USA*, 101: 4966-4971.

L

LEE Genee Y., KENNY Paraic A., LEE Eva H. & BISSELL Mina J. (2007), 'Three-dimensional culture models of normal and malignant breast epithelial cells', *Nature Methods*, 4: 359-365.

LIU Edison T., KUZNETSOV Vladimir A. & MILLER Lance D. (2006), 'In the pursuit of complexity: systems medicine in cancer biology', *Cancer Cell*, 9: 245-247.

M

MAFFINI Maricel V., SOTO Ana M., CALABRO Janine M., UCCI Angelo A. & SONNENSCHNEIN Carlos (2004), 'The stroma as a crucial target in rat mammary gland carcinogenesis', *Journal of Cell Science*, 117: 1495-1502.

MALATERRE Christophe (2007), 'Organicism and reductionism in cancer research: towards a systemic approach', *International Studies in the Philosophy of Science*, 21: 57-73.

- MARCUM James A. (2005), 'Metaphysical presuppositions and scientific practices : reductionism and organicism in cancer research', *International Studies in the Philosophy of Science*, 19: 31-45.
- MARCUM James A. (2009), *The conceptual foundations of systems biology: an introduction*, New York, Nova Scientific Publishers.
- MARCUM James A. & VERSCHUUREN Geert M.N. (1986), 'Hemostatic regulation and Whitehead's philosophy of organism', *Acta Biotheoretica*, 35: 123-133.
- MCDANIEL Shauntae M., RUMER Kristen K., BIROC Sandra L., METZ Richard P., SINGH Meenakshi, PORTER Weston & SCHEDIN Pepper (2006), 'Remodeling of the mammary microenvironment after lactation promotes breast tumor cell metastasis', *American Journal of Pathology*, 168: 608-620.
- MEDINA Daniel (2004), 'Stromal fibroblasts influence human mammary epithelial cell morphogenesis', *Proceedings of the National Academy of Science USA*, 101: 4723-4724.
- MEDINA Daniel & KITTRELL Frances (2005), 'Stroma is not a major target in DMBA-mediated tumorigenesis of mouse mammary preneoplasia', *Journal of Cell Science*, 118: 123-127.
- MESAROVIC Mihajlo D., SREENATH Sree N. & KEENE Jack D. (2004), 'Search for organizing principles: understanding in systems biology', *Systems Biology*, 1: 19-27.
- MORANGE Michel (1993), 'The discovery of cellular oncogenes', *History and Philosophy of the Life Sciences*, 15: 45-58.
- MOSS Lenny (2003), *What genes can't do*, Cambridge, MIT Press.
- N**
- NELSON Celeste M. & BISSELL Mina J. (2005), 'Modeling dynamic reciprocity: engineering three dimensional culture models of breast architecture, function, and neoplastic transformation', *Seminars in Cancer Biology*, 15: 342-352.
- NOBLE Denis (2003), 'The future: putting Humpty-Dumpty together again', *Biochemical Society Transactions*, 31: 156-158.
- O**
- ORIMO Akira, GUPTA Piyush B., SGROI Dennis C., ARENZANA-SEISDEDOS Fernando, DELAUNAY Thierry, NAEEM Rizwan, CAREY Vincent J., RICHARDSON Andrea L. & WEINBERG Robert A. (2005), 'Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion', *Cell*, 121: 335-348.
- P**
- PLACEK Tomasz (2004), 'Quantum state holism: a case for holistic causation', *Studies in the History and Philosophy of Modern Physics*, 35: 671-692.
- PROSTATE CANCER COMPLEXITY WORKING GROUP (2002), 'Merging bottom-up and top-down approaches to study prostate cancer biology', *Complexity*, 7: 22-30.
- R**
- ROOT-BERNSTEIN Robert S. (1999), 'Complementarity and contradiction in cancer research: the role of hierarchies in carcinogenesis', *Anticancer Research*, 19: 4915-4918.
- ROTHMAN Kenneth J. (1974), 'Synergy and antagonism in cause-effect relationships', *American Journal of Epidemiology*, 99: 385-388.
- S**
- SADLER T.M., GAVRIIL M., ANNABLE T., FROST P., GREENBERGER L.M. & ZHANG Y. (2006), 'Combination therapy for treating breast cancer using antiestrogen, ERA-923, and the mammalian target of rapamycin inhibitor, temsirolimus', *Endocrine Related Cancer*, 13: 863-873.
- SCHMEICHEL Karen L. & BISSELL Mina J. (2003), 'Modeling tissue-specific signaling and organ function in three dimensions', *Journal of Cell Science*, 116: 2377-2388.
- SCOTT Alwyn C. (2007), *The non-linear universe: chaos, emergence, life*, New York, Springer.

- SONNENSCHNEIN Carlos & SOTO Ana M. (1999), *The society of cells: cancer and control of cell proliferation*, New York, Springer.
- SONNENSCHNEIN Carlos & SOTO Ana M. (2005), 'Are times a "changing" in carcinogenesis?', *Endocrinology*, 146: 11-12.
- SONNENSCHNEIN Carlos & SOTO Ana M. (2006), 'Carcinogenesis and metastasis now in the third dimension – what's in it for pathologists?', *American Journal of Pathology*, 168: 363-366.
- SOTO Ana M. & SONNENSCHNEIN Carlos (2004), 'The somatic mutation theory of cancer: growing problems with the paradigm?', *BioEssays*, 26: 1097-1107.
- SOTO Ana M. & SONNENSCHNEIN Carlos (2005a), 'Response to Coffman', *BioEssays*, 27: 461.
- SOTO Ana M. & SONNENSCHNEIN Carlos (2005b), 'Emergentism as a default: cancer as a problem of tissue organization', *Journal of Biosciences*, 30: 103-118.
- SOTO Ana M. & SONNENSCHNEIN Carlos (2006), 'Emergentism by default: a view from the bench', *Synthese*, 151: 361-376.
- SPENCER Virginia A., XU Ren & BISSELL Mina J. (2007), 'Extracellular matrix, nuclear and chromatin structure, and gene expression in normal tissues and malignant tumors: a work in progress', *Advances in Cancer Research*, 97: 275-294.
- SPORN Michael B. (2006), 'Dichotomies in cancer research: some suggestions for a new synthesis', *Nature Clinical Practice Oncology*, 3: 364-373.
- SUTER Rachel & MARCUM James A. (2007), 'The molecular genetics of breast cancer and targeted therapy', *Biologics: Targets & Therapy*, 1: 243-257.
- T**
- TREWAVAS Anthony (2006), 'A brief history of systems biology', *The Plant Cell*, 18: 2420-2430.
- V**
- VAN DE VIJVER Gertrudis, VAN SPEYBROECK Linda & VANDEVYVERE Windy (2003), 'Reflecting on complexity of biological systems: Kant and beyond?', *Acta Biotheoretica*, 51: 101-140.
- VAN REGENMORTEL Marc H. (2004), 'Biological complexity emerges from the ashes of genetic reductionism', *Journal of Molecular Recognition*, 17: 145-148.
- W**
- WANG Fei, WEAVER Valerie M., PETERSEN Ole W., LARABELL Carolyn A., DEDHAR Shoukat, BRIAND Per, LUPU Ruth & BISSELL Mina J. (1998), 'Reciprocal interactions between beta1-integrin and epidermal growth factor receptor in three-dimensional basement membrane breast cultures: a different perspective in epithelial biology', *Proceedings of the National Academy of Science USA*, 95: 14821-14826.
- WANG Yong, THAKUR Archana, SUN Yuan, WU Jack, BILIRAN Hector & BOLLIG Alicia (2007), 'Synergistic effect of cyclin D1 and c-Myc leads to more aggressive and invasive mammary tumors in severe combined immunodeficient mice', *Cancer Research*, 67: 3698-3707.
- WEAVER Valerie M. & GILBERT Penney (2004), 'Watch they neighbor: cancer is a communal affair', *Journal of Cell Science*, 117: 1287-1290.
- WEINBERG Robert A. (1988a), 'The genetic origins of human cancer', *Cancer*, 61: 1963-1968.
- WEINBERG Robert A. (1988b), 'Finding the anti-oncogene', *Scientific American*, 259: 44-51.
- WEINBERG Robert A. (1998), *One renegade cell: How cancer begins*, New York, Basic Books.

◉The Author◉

James A. Marcum is professor of philosophy and director of the Medical Humanities Program at Baylor University in Waco, Texas, USA. He earned doctorates in philosophy from Boston College and in physiology from the University of Cincinnati Medical College. He was a postdoctoral fellow at Massachusetts Institute of Technology and a faculty member at Harvard Medical School for over a decade before coming to Baylor University. He received grants from several funding agencies, including the National Institutes of Health and the National Science Foundation. His current research interests include the philosophy and history of science and medicine. Examples of his recent publications appear in *Annals of Science*, *International Studies in the Philosophy of Science*, *Synthese*, *Perspectives on Science*, *History and Philosophy of the Life Sciences*, *Journal of the History of Medicine and Allied Sciences*, *Medicine, Health Care and Philosophy*, *Perspectives in Biology and Medicine*, and *Theoretical Medicine and Bioethics*. His most recent books are *An introductory philosophy of medicine: humanizing modern medicine*. Philosophy and Medicine Series, volume 99, New York, Springer, 2008, and *The conceptual foundations of systems biology: an introduction*. Systems Biology-Theory, Techniques and Application Series, Hauppauge, NY, Nova Science Publishers, 2009.

Department of Philosophy, One Bear Place #97273, Baylor University, Waco, TX 76798 USA, 254-710-3745 (office).

James_Marcum@baylor.edu

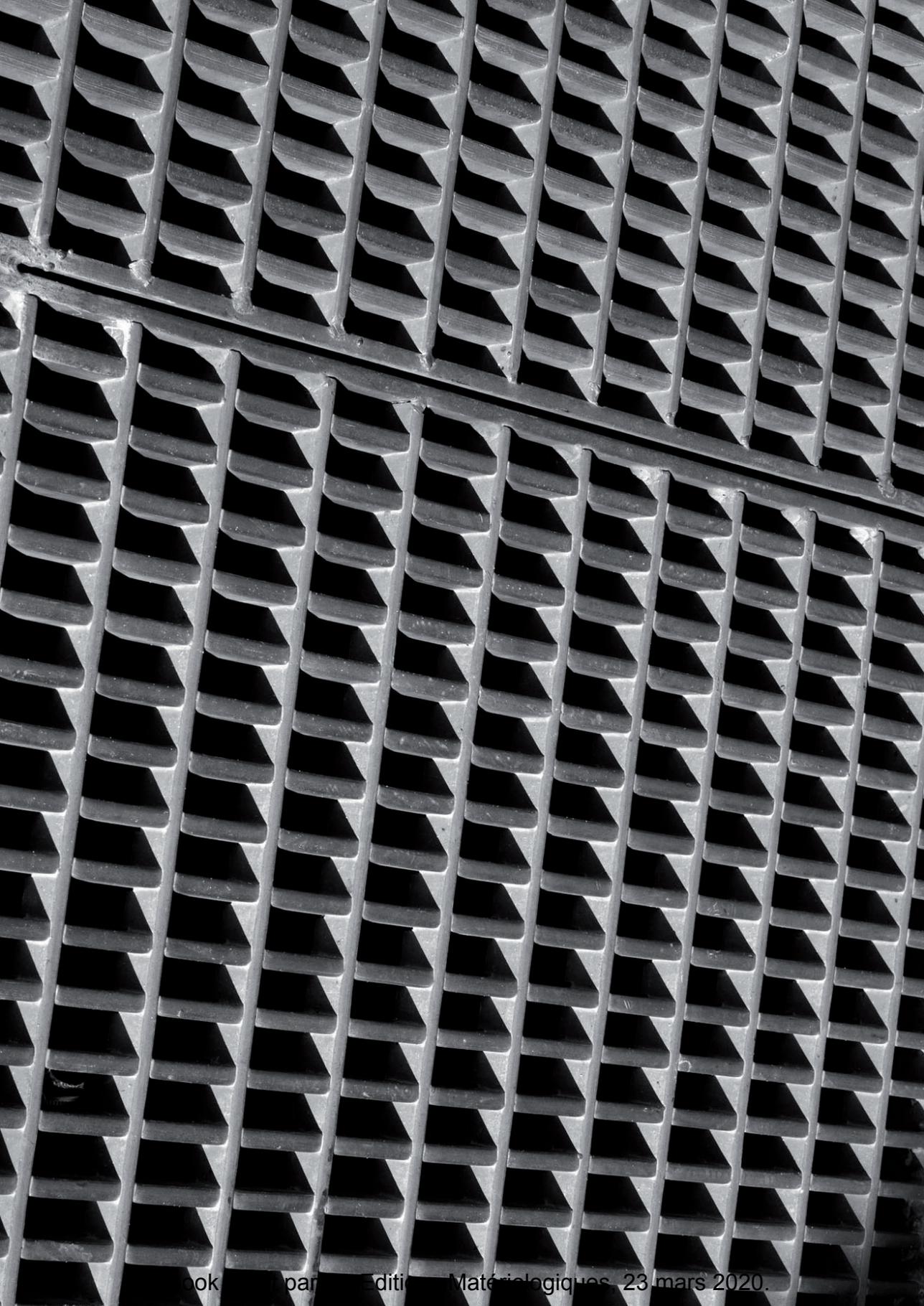
Avertissement. Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur et des auteurs. Les textes et illustrations figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage personnel, scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document selon ce modèle :

James A. Marcum, «Cancer: complexity, causation, and systems biology», in Gérard Lambert et Marc Silberstein (dir.), Matière première. Revue d'épistémologie [en ligne]. Nouvelle série, N° 1/2010 : Épistémologie de la médecine et de la santé, Éditions Matériologiques. Mis en ligne le 7 octobre 2010. URL : www.materiologiques.com

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.

**T
A
B
A
C**

11^{me}. ARR^t.
**RUE
DE LA
FONTAINE AU ROI**



**Bernard Swynghedauw
& Christian Frelin** **L'évolution
biologique en médecine.**
**Un outil de formation, un moyen de
hiérarchiser les informations et une
base rationnelle à une politique de santé**

Évolution biologique et médecine ne sont pas des compagnons de route habituels. L'enseignement de la biologie de l'évolution au début des cycles de biologie dans les facultés de médecine est presque inexistant en France comme dans la plupart des pays (NESSE *et al.* 2009). Cette branche essentielle de la biologie n'est pas familière au corps médical (encadré 1), pour le médecin l'évolution est encore du domaine de la philosophie, et Darwin est tout sauf un simple biologiste (on disait de son temps un naturaliste) génial qui a marqué la discipline de façon indélébile. Sur un plan médical, l'un des développements les plus importants de la biologie de l'évolution concerne la politique de santé. La médecine évolutionniste est un concept généré par un psychiatre, Randolph Nesse il y a une quinzaine d'années. Ce concept est un peu devenu la propriété de cercles particuliers peu concernés par la médecine de tous les jours (NESSE & WILLIAMS 1994, TREVATHAN *et al.* 2007, STEARNS & KOELLA 2007, GLUCKMAN *et al.* 2009, ZIMMER 2010). Ce concept, sur lequel nous avons déjà publié (SWYNGHEDAUF 2009), servira d'illustration.

On propose ici que l'évolution biologique soit introduite dans l'enseignement médical, vétérinaire et pharmaceutique, et de délimiter

à cet effet quelques-uns des obstacles qui en rendent l'application si difficile. Les informations biologiques n'ont de sens que dans le cadre de l'évolution et recadrer le fait médical¹ à ce niveau est un moyen simple de mieux enseigner la biologie et, pour les chercheurs en biomédecine, de hiérarchiser le torrent d'informations qui submerge actuellement le corps médical. Notre génome contemporain a été façonné durant des milliards d'années par un environnement instable. Le fait nouveau, jamais expérimenté jusqu'alors, c'est que l'environnement, notre environnement, a brutalement et radicalement changé, en bien comme en mal. Il y a moins de famines, on vit mieux et plus vieux, mais en revanche nos ressources en énergies fossiles sont en train de s'épuiser (JANCOVICI 2009) et l'homme est en train de modifier le climat, ce qui peut avoir des conséquences graves imprévisibles (JOUZEL & DEBROISE 2007). Le paysage médical s'est également complètement modifié, reflétant un conflit génotype/environnement d'un type nouveau (Swynghedauw 2009).

Cette manière de voir permet d'isoler un certain nombre de mécanismes essentiels et très anciens. La démarche est potentiellement riche en conséquences aussi bien pour la pratique médicale que pour la recherche pharmacologique. Les exemples qui seront donnés ont été choisis dans la pathologie courante, quotidienne, celle qui est responsable de l'essentiel de la mortalité et de la morbidité au XXI^e siècle dans nos pays. Finalement et à ce titre, cette démarche pourrait aussi être considérée

1. Peut-être faut-il, à propos de «fait médical», rappeler cette vérité essentielle, à savoir que la médecine est une science appliquée dont le seul objet est la prévention et le traitement des maladies (il ne peut d'ailleurs s'agir que de simples symptômes sans étiquettes précises), qu'elles soient réelles, simulées ou simplement ressenties. Pour le médecin, tout ce qui guérit est bon, mais encore faut-il qu'il s'agisse bien d'un effet thérapeutique. L'argument dit de «la grand-mère», si cher aux partisans de la médecine douce et aux homéopathes – «Docteur, ma grand-mère est morte à 80 ans, elle fumait le cigare, avait un cancer et prenait tous les jours de l'eau de mélisse» –, n'est pas un argument. Il reste en effet à démontrer que (i) le diagnostic avait bien été fait et que (ii) ladite eau de mélisse est plus efficace que rien du tout ! L'Académie nationale de médecine, dans un de ses rapports récents, a fait, courageusement et à contre-courant des idées reçues, un sort à ce genre de pratique.

Encadré 1. Les difficultés rencontrées dans l'enseignement de la biologie de l'évolution aux futurs médecins

Intégrer ces données basiques de l'évolution est, pour le médecin, une démarche presque contre nature.

(1) L'évolution n'a aucune finalité. Une vue anthropocentrique de l'évolution est une vue faussée à la base ; à l'opposé, le médecin, par métier, est anthropocentrique, un médecin qui n'aurait pas pour base l'homme serait un mauvais médecin. Replacer l'homme dans l'histoire du vivant est donc, pour le médecin, une démarche qui va à l'encontre de son éthique personnelle et de sa pratique quotidienne.

(2) La biologie de l'évolution concerne des populations et pas des individus. À l'opposé, pour le médecin, le cadre morbide, la maladie, n'est qu'une manière provisoire d'asseoir son diagnostic et d'établir un traitement. Son malade est toujours unique et plus compliqué. Dans le meilleur des cas, il s'agit de retrouver d'autres malades – au besoin au moyen de modèles animaux en médecine expérimentale – ayant les mêmes signes cliniques, en sachant que ces signes évoluent avec la technologie, ils ne sont pas les mêmes en ces temps de scintigraphie qu'ils l'étaient du temps de Laennec et du stéthoscope. L'histoire de la médecine est remplie de cadres morbides remaniés (comme c'est le cas en ce moment pour l'asthme (WENZEL 2006) ou le cancer (STRATTON *et al.* 2009)).

(3) La génétique a modifié la pratique médicale. Dans le cas des maladies monogéniques, la relation entre la mutation causale et le phénotype (qui est le tableau clinique) est claire. Les maladies monogéniques sont cependant très rares. La génétique médicale est aujourd'hui confrontée au problème posé par les maladies dites communes. La relation génotype/phénotype est ici infiniment plus complexe : plusieurs gènes peuvent contribuer au trait, et un même gène peut influencer plusieurs traits en même temps, on parle de déterminisme multigénique ; interviennent à la fois des interactions intergéniques et des modifications de l'expression génique dues à l'environnement et véhiculées par les mécanismes épigénétiques.

(4) La variabilité du génome humain est bien établie. Il n'existe pas d'individus génétiquement identiques (si l'on excepte les vrais jumeaux). La composition génétique des populations humaines se modifie avec le temps, mais cette évolution est lente, beaucoup plus lente que la longévité du praticien et de ses patients.

(5) L'adaptation, qui est l'appropriation d'un organisme aux conditions de l'existence, permet à l'individu de durer et de se reproduire. Il n'y a pas plusieurs sens à ce mot, mais simplement ce terme n'est pas toujours utilisé dans le même contexte. Pour le biologiste de l'évolution, l'adaptation se situe sur une échelle de temps de l'ordre du million d'années, temps nécessaire à l'apparition et à la sélection d'un gène codant une protéine et une fonction utiles.

Par contre, l'adaptation physiologique (comme l'adaptation à l'effort ou à la déshydratation), qui est familière au médecin, s'observe sur des durées de l'ordre de la seconde ou de l'heure, voire de la journée. Le mécanisme des processus d'adaptation est en général commun à toutes les espèces. C'est le mécanisme qui est transmissible. L'exemple le plus connu est celui de la mécanotransduction présente chez les arbres, les bactéries, comme chez l'homme où elle rend compte de l'hypertrophie cardiaque des hypertendus (LOMPRE 1979, SWYNGHEDAUF 1999) comme celle des athlètes (SWYNGHEDAUF 1986).

comme un des fondements rationnels d'une politique de santé au long court (SWYNGHEDAUF 2009).

Darwin n'était pas médecin, mais, sous la pression de son père, il avait commencé sa médecine qu'il avait rapidement abandonnée, après deux ans, dégoûté qu'il était par la manière assez rude dont on pratiquait la chirurgie sans anesthésie à son époque, en sortant de la salle d'autopsie. Le métier de médecin est, par définition, anthropocentrique, l'évolution ne l'est en aucune façon et il est important de rappeler cette contradiction essentielle.

1 ➔ Enseigner la biologie de l'évolution à de futurs médecins. Comment ?

La biologie de l'évolution comporte des aspects purement descriptifs, à la fois horizontaux, la vie telle qu'on peut l'observer aujourd'hui, et des aspects verticaux, historiques ; la vie existe depuis environ 3,6 milliards d'années, et durant cette période de nombreuses espèces sont apparues ou ont disparu. Il existe par ailleurs des aspects mécanistiques que l'on peut résumer avec la formule suivante : la diversité (génétique) précède la sélection et la spéciation. Enseigner la biologie de l'évolution comporte donc quatre volets distincts, certains d'entre eux sont bien évidemment incontournables (ils sont résumés dans l'encadré 2), mais, dans le détail, certains de ces aspects sont d'intérêt inégal pour le médecin et enseigner cette discipline à de futurs médecins ou à des médecins en formation permanente oblige à des choix et comporte des priorités qui ne sont pas toujours celles du zoologiste ou du biologiste.

1.1 ➔ L'histoire de la vie

La vie et les principales cascades métaboliques ont une très longue histoire, dont la connaissance a un certain nombre de conséquences pour la pratique, ou plus précisément pour la conceptualisation du fait médical. Donnons-en quelques exemples :

Encadré 2. Quelques notions de base concernant l'évolution biologique

(1) Le « dogme central de la biologie moléculaire », que l'on doit à Francis Crick, est le suivant : le code génétique dont l'ADN, l'acide désoxyribonucléique, est le support chimique, est transcrit en ARN, acide ribonucléique (messenger), puis traduit en protéines. Ce sont les protéines qui déterminent la fonction et le phénotype. Cette réaction est irréversible. Il y a quelques rares exceptions, lesquelles n'ont aucune conséquence majeure en biologie.

(2) Le processus évolutif n'est possible que s'il y a des mutations, c'est-à-dire des changements dans l'arrangement des bases composant le génome, mais il faut savoir que le phénotype lui-même peut être modifié autrement sous l'influence de plusieurs mécanismes, les plus importants étant de nature épigénétique.

(3) La notion de temps est majeure, la vie a débuté il y a plusieurs milliards d'années. Évolution, par définition, signifie la vie en fonction du temps, l'évolution est une branche de l'histoire, mais la constante de temps est ici de l'ordre du million d'années, et n'a rien à voir avec la constante de temps à laquelle les historiens nous ont habitués.

(4) Le moteur de l'évolution c'est la pression sélective exercée par l'environnement qui assure un succès reproductif différentiel (c'est la définition de la fitness) et donc la sélection, naturelle ou artificielle (la sélection artificielle pratiquée par les éleveurs repose à la fois sur le choix fait par l'homme de propriétés utiles, l'apparition spontanée de variations et l'utilisation d'un grand nombre d'individus pour cette sélection, c'est le premier chapitre de l'*Évolution des espèces* de Darwin). La sélection n'a donc rien d'aléatoire, si l'on excepte bien entendu les variations de l'environnement qui, elles, le sont.

(5) Le rôle du hasard est très important. Il détermine la variabilité des individus au sein d'une population ; le hasard précède la sélection et la spéciation, lesquelles se font à la fois par pression sélective darwinienne (c'est-à-dire par nécessité) et par le simple hasard de la dérive génétique. On peut reconstruire notre histoire, mais on ne peut prédire ni le futur d'une espèce, ni celui d'un individu car ils dépendent de phénomènes totalement imprévisibles.

- Le premier concerne la naissance des premières molécules organiques et de la première cellule. Le code génétique, ou, de façon plus générale, les mécanismes assurant la transcription du support chimique de l'hérédité, à savoir le code génétique présent dans l'ADN, et sa traduction en protéines, sont les mêmes chez l'homme et chez les bactéries. Cette donnée constitue le plus simple des arguments à opposer à ceux qui nient l'existence du phénomène évolutif, c'est aussi la base d'une des disciplines de base de la recherche biomédicale, la microbiologie. Ce qui est vrai pour *Escherichia coli* est aussi vrai pour l'éléphant, a dit un prix Nobel.

- Le second concerne la mort, la mort cellulaire. La mort cellulaire peut survenir après un coup violent ou en l'absence d'oxygène, mais elle peut aussi être programmée par la cellule elle-même, c'est la mort cellulaire par apoptose. L'apoptose est aussi ancienne que la vie, elle existe chez l'homme comme chez la bactérie (AMEISEN 2002), et, pour simplifier, on peut dire que le développement embryonnaire n'est possible que si disparaissent certaines cellules pour qu'en apparaissent de nouvelles (l'apoptose comme «*eat-me-signal*»). Le médecin connaît de nombreuses maladies impliquant l'apoptose, le cancer qui a perdu cette propriété (HANAHAN & WEINBERG 2000), mais à l'inverse l'insuffisance cardiaque dans laquelle l'activation de l'apoptose joue un rôle aggravant (GARG *et al.* 2004).

- La survie, pour toutes les espèces vivantes, dépend de la capacité à réparer les dommages causés par les diverses agressions dont nous sommes en permanence l'objet. Le système de réparation, qui inclut le processus inflammatoire, le système immunitaire, la coagulation et la cicatrisation, est, également, aussi ancien que la vie. Ce système est par contre assez différent au cours de l'évolution. Le système immunitaire a deux composantes, interdépendantes, le système inné, transmissible, et le système immunitaire acquis, dit aussi adaptatif. Le premier existe depuis toujours, le second n'existe pas chez les plantes (NÜMBERGER *et al.* 2004), il est donc apparu à une date précise de l'évolution, celle où sont apparues les premières plantes. L'histoire des protéases qui permettent la coagulation est également informative (DAVIDSON *et al.* 2003).

1.2 ↻ La classification du vivant (ou systématique)

La nouvelle systématique, dite hennigienne, du nom de son inventeur, Willi Hennig, ou phylogénétique, classe le vivant en clades ou taxons monophylétiques. La méthode complexe caractérise la séquence évolutive des transformations selon les principes découverts par Darwin (LECOINTRE 2001).

Souris, rat et lapin, qui sont les espèces animales habituellement utilisées en médecine expérimentale, sont également, sur un plan phylogénétique, proches des primates². C'est la raison pour laquelle bien

2. Les lapins qui sont des lagomorphes et les souris et les rats qui sont des rongeurs appartiennent tous deux aux glires, sous-groupe des euarchontoglires et des mammifères placentaires, tout comme les primates.

des mécanismes physiopathologiques nouveaux souvent très subtils existent aussi bien chez l'homme que chez les rongeurs. Par exemple, le mécanisme par lequel les microARN³ (SMALL *et al.* 2010) régulent l'expression des gènes dans l'insuffisance cardiaque est le même chez l'homme et chez la souris.

La systématique phylogénétique a révolutionné la biologie, mais son intérêt médical est plus limité. Elle a permis l'étude des pandémies virales en facilitant l'établissement des filiations des souches virales ou bactériennes. Des événements récents ont montré le caractère hautement imprévisible des mutations et l'impossibilité où l'on est de prédire l'apparition des souches pathogènes.

1.3 ↻ La diversité (ou polymorphisme) génétique

Le polymorphisme génétique, condition indispensable à la sélection et à la génétique moléculaire, est à l'origine de la diversité des populations. Cette diversité est à la fois le vivier dans lequel l'évolution pêche de nouveaux génotypes et le responsable de nos apparences ou de nos capacités qui font que nous sommes tous différents dans nos aspects physiques et comportementaux. Une définition récente du gène consisterait à considérer le gène non pas dans sa nature physique – un ou des segments particuliers de la chaîne ADN capables de transcrire puis de traduire une ou des protéines – mais dans sa fonction, les gènes n'étant que la capacité que chacun a de réagir à l'environnement. Finalement, le gène, le génome, n'ont aucun sens s'ils ne servent à rien, et leur seule fonction est de réagir à l'environnement (GERSTEIN *et al.* 2007)⁴. Cette définition inclut surtout le fait que nous réagissons tous d'une façon

3. Les microARN sont des ARN de très petite taille qui sont transcrits mais qui ne sont pas traduits en protéines. Ils peuvent s'hybrider avec les séquences correspondantes des ARN messagers, ceux qui servent de copie pour la traduction, et bloquer la traduction.

4. Quelques auteurs, au nom d'un humanisme aux assises essentiellement métaphysiques et en se basant à la fois sur la faible productivité des enquêtes génétiques et sur la complexité du comportement et des maladies du comportement, se sont crus autorisés à remettre en cause le paradigme central de la vie, émis par Crick, à savoir que l'ADN est à l'origine des protéines et que celles-ci sont à l'origine des fonctions. Rien n'autorise sérieusement. une telle remise en cause qui, d'une manière générale est la négation de la méthode scientifique laquelle procède patiemment, par erreurs et succès

très spécifique aux modifications environnementales, la température externe par exemple – certains sont plus frileux que d'autres – mais aussi le régime «cafétéria» – tout le monde n'est pas obèse. Le nombre d'exemples est infini. Cette nouvelle manière de concevoir le gène repose sur les dernières analyses du programme ENCODE (ce programme est en plein développement et ses résultats peuvent facilement être trouvés sur Internet, l'article de Gerstein en résume quelques-uns des résultats importants) et la découverte, entre autres, de gènes dont les exons ou les parties régulatrices sont répartis dans plusieurs chromosomes.

Cette réactivité, qui est propre à tout un chacun, porte également sur la manière que nous avons de réagir face aux aspects agressifs de notre environnement, les infections, les excès alimentaires, les toxiques cancérigènes, les radiations ionisantes... Tout le monde ne fait pas une tuberculose généralisée, un infarctus, un diabète ou un cancer. Il existe, rarement, on le sait et on sait pourquoi, des tuberculoses, des lèpres, etc., familiales (certains individus n'ont pas les gènes de l'immunité qui permettent de se défendre contre ce type d'infection)⁵. Il existe aussi – c'est plus fréquent – des infarctus du myocarde ou des diabètes, ou – c'est plus rare –, des cancers familiaux... Le facteur héréditaire est un risque, mais un risque variable. La variabilité du risque a à la fois un aspect populationnel, ce type de risque n'est pas le même pour toutes les maladies (le risque héréditaire est élevé pour le diabète, relativement faible pour l'hypertension artérielle), et un aspect individuel, le risque dépend de l'existence d'anomalies génétiques lesquelles sont inégalement réparties d'une personne à l'autre.

La maladie est clairement le résultat d'un conflit entre un génome, qui est notre propriété privée, et un environnement, qui nous est tous commun, ou presque (SWYNGHEDAUW 2009). Par manque d'outil approprié, la recherche génétique a pendant longtemps porté sur les affections héréditaires dites monogéniques comme la mucoviscidose par exemple. Ce type d'affection génétique est rare et de mécanisme relativement

progressifs. C'est aussi faire fi des données nouvelles fournies par les nouvelles orientations de la biologie modulaire et spatiale, et l'épigénétique.

5. Les tuberculoses familiales sont très rares et bien évidemment, ce n'est pas ce paramètre génétique qui explique la prédominance de cette affection dans nos pays jusqu'en 1946, la misère, la guerre, mais aussi l'absence de traitement efficace bon marché ont été les facteurs déterminants.

simple, l'hérédité de l'affection est forte et familiale, une mutation génétique précise perturbe une fonction comme celle d'un canal ionique (le canal chlorure, CFTR) spécifique dans la mucoviscidose (KAPLAN & DELPECH 2007). À l'opposé, il y a de multiples affections qui n'ont rien d'héréditaire, l'accident de la route, une infection massive, le simple rhume de cerveau.

Le vrai problème concerne les maladies graves courantes, le quotidien du généraliste, comme l'infarctus du myocarde par exemple; une notion familiale y est un facteur de risque important et un des éléments du diagnostic. C'est dans ce type d'affection qu'il faut chercher non pas une origine génétique unique, mais des facteurs génétiques multiples dits favorisants. La génétique a récemment développé une nouvelle approche permettant l'analyse du génome entier (*genomewide analysis*, GWA⁶). Ces techniques identifient des secteurs du génome, des loci, où certains polymorphismes sont plus fréquents dans une population de malades que chez des sujets normaux. Certains de ces *loci* contiennent des gènes qui deviennent des coupables potentiels, ce qui crée de nouvelles pistes de recherche aussi bien diagnostiques que thérapeutiques.

La quasi-totalité des affections, y compris les affections plus complexes comme les maladies psychiatriques par exemple, ont maintenant fait l'objet de cette nouvelle approche (DONNELLY 2008). Dans l'ensemble, chaque affection a son profil particulier, et le nombre de *loci* y est généralement important ce qui, pour le moment, laisse à penser que ces affections sont en fait des syndromes regroupant des maladies différentes; certains vont même jusqu'à dire qu'il y a autant de maladies que de malades. Ce concept est à la base de la médecine dite personnalisée. Cette avancée technologique ne signe néanmoins pas la fin de l'histoire car il apparaît de plus en plus évident que l'identification de gènes coupables obtenue par ce procédé ne met finalement en évidence qu'un nombre limité d'associations simples, suggérant l'existence de modifi-

6. La technique repose sur deux principes: (i) l'analyse systématique de l'ensemble du génome au moyen de sondes détectant des polymorphismes ponctuels à intervalles plus ou moins réguliers tout le long du génome, (ii) l'utilisation d'une carte des haplotypes, c'est-à-dire des regroupements de polymorphisme ponctuels; cette carte permet d'élargir le champ de l'exploration.

cations plus complexes qui sortent un peu du cadre de cet article (voir par exemple la «chronique génomique» de Bertrand JORDAN 2010).

Un exemple. L'infarctus du myocarde est la manifestation clinique la plus connue de l'athérosclérose artérielle. C'est une maladie complexe multifactorielle dont le risque est accru à la fois par l'existence d'antécédents familiaux et par de nombreux facteurs environnementaux comme la surcharge pondérale, le tabac, la sédentarité, l'âge... L'analyse génétique par GWA a mis en évidence de façon concordante dans plusieurs laboratoires de nombreux *loci* liés à la maladie. Certains de ces *loci*, les plus rares, contiennent des gènes à risque, c'est-à-dire des gènes potentiellement capables de rendre compte du développement de l'affection; d'autres, les plus fréquents, tout à fait inattendus, ont une fonction apparemment sans lien avec l'affection mais ils ont ouvert des pistes de recherche nouvelles, bien nécessaires dans l'état actuel de la recherche sur ce type d'affection (McPHERSON 2010).

La résistance aux antibiotiques, aux antifongiques, aux antipaludiques surtout, est un problème d'une portée croissante considérable dans tous les pays du monde. Les mécanismes à l'origine de cette résistance sont multiples et essentiellement génétiques. Ce chapitre a des retombées financières considérables. Ce que l'on n'hésite pas à qualifier de «*malaria disaster*» questionne toutes les grandes organisations mondiales, d'autant que la très grande majorité des malades n'étant pas solvables, les fonds nécessaires à la recherche de nouveaux antipaludiques ne peuvent venir que de contributions publiques et non commerciales, lesquelles tardent à se concrétiser. Le problème tient probablement plus au développement qu'à la recherche, car il existe des milliers d'antipaludéens dans le pipeline (le sujet fait l'objet d'une récente mise au point dans *Nature*, FIDOCK 2010)!

L'étude de la résistance aux antipaludéens de synthèse a fourni de bons exemples des applications de la biologie de l'évolution à la thérapeutique. Plusieurs erreurs ont ainsi pu être identifiées dans ce domaine (READ & HUIJBEN 2009). Contrairement à l'opinion répandue, (i) les antipaludéens actifs sur la transmission du parasite (les gamétocides) ne ralentissent pas la diffusion de la résistance; (ii) les antipaludéens à longue vie ne sont pas plus efficaces que les autres; (iii) les mutations *de novo* ne sont pas le principal responsable de la résistance; (iv) les parasites ne sont probablement pas fragilisés lorsqu'ils deviennent résistants; (v) la fixation de la résistance n'est pas inévitable si la

pression exercée par le traitement est maintenue. Ce type d'approche évolutionniste est ici clairement utile pour rendre efficace une politique de santé de grande portée.

1.4 ➔ La sélection des caractères et la spéciation

Au cours de l'histoire de la vie, certains caractères ou comportements sont sélectionnés par hasard ou par nécessité (MONOD 1970). La nécessité, c'est le déterminisme, le déterminisme darwinien, c'est elle qui constitue la force qui pousse l'évolution. Le hasard, c'est le hasard de la dérive génétique, mais aussi le simple hasard qui guide les variations de l'environnement.

La spéciation, c'est la survenue de nouvelles espèces vivantes qui, par définition, ne se reproduiront pas entre elles, créant ainsi *de facto* une barrière dans la reproduction. La spéciation, c'est-à-dire «l'apparition de nouvelles barrières de reproduction» (LECOINTRE 2009), est un phénomène continu qui a toujours lieu. La survenue d'une nouvelle espèce a longtemps été attribuée à des changements permettant une meilleure adaptation à un nouvel environnement. Des analyses plus précises ont montré que la réalité était infiniment plus compliquée. Des fusions chromosomiques ont par exemple pu survenir, par hasard, au cours de l'isolement géographique, et rendre ainsi les nouvelles espèces incapables de se reproduire les unes avec les autres ; c'est l'exemple, maintenant classique, d'espèces de souris importées dans l'île de Madère ou d'espèces de moustiques dispersées dans les couloirs du métro londonien (LECOINTRE 2009). Plus récemment, une analyse extensive (portant sur les levures, la drosophile, la souris, les plantes) a confirmé à la fois la multiplicité des mécanismes impliqués et le fait que les gènes dits de spéciation – les premiers à apparaître dans la nouvelle lignée – ne sont habituellement pas adaptatifs (PRESGRAVES 2010).

L'apparition de nouvelles maladies génétiques procède du même mécanisme, mais là aussi hasard et nécessité forment un tout inextricable. Les gènes délétères responsables des maladies monogéniques ont une histoire souvent connue et sont apparus à une date relativement précise par pur hasard (JOBBLINGS *et al.* 2004). Certains, relativement bénins, sont conservés par pur hasard et se sont maintenus également par hasard par le seul jeu de la dérive génétique. D'autres, apparus également par hasard, se sont maintenus parce qu'ils pouvaient donner un avantage aux porteurs dans un environnement spécifique. L'exemple le plus connu est celui de la mutation du gène de l'hémoglobine qui est

à l'origine de l'anémie falciforme (*sickle cell disease* ou drépanocytose) ainsi nommée car la mutation fragilise l'hémoglobine et déforme les globules rouges. Cette affection redoutable est extrêmement répandue dans les populations d'origine africaine. Elle est responsable d'une anémie grave. Les patients homozygotes meurent d'anurie (perte de la fonction rénale) vers la quarantaine. Cette mutation a néanmoins été conservée et sélectionnée et tend même à se répandre en Afrique car elle protège les porteurs contre le paludisme. Le paludisme est l'une des maladies les plus fréquentes du monde, c'est l'un des premiers responsables de la mortalité infantile africaine. *Plasmodium falciparum*, le parasite responsable de cette affection, se nourrit d'hémoglobine et la présence de la mutation (chez les hétérozygotes) crée un environnement dans lequel le parasite se reproduit mal (ceci a été confirmé par des études *in vitro*). Les porteurs de la mutation sont ainsi protégés contre la cause de mortalité la plus importante dans ces pays, et vivent plus longtemps que les non-porteurs.

Les exemples de ce type sont rares bien qu'ils touchent des millions de personnes. L'histoire de l'obésité a plus de portée dans nos pays. L'obésité n'est pas qu'un problème esthétique, c'est aussi un risque considérable de maladies cardiovasculaires, mais aussi de cancer et de bien d'autres affections courantes. L'obésité est à la fois due à des facteurs environnementaux – il n'y a pas d'obésité dans les périodes de famine – et à des facteurs génétiques – l'héritabilité λ est élevée (autour de 60 %) et la GWA a identifié près de 50 anomalies génétiques associées à l'obésité. Ces mutations sont le fruit d'une sélection qui date, en partie, de l'époque où la nourriture était rare et irrégulière, le hasard a fait apparaître des gènes permettant le stockage des graisses plus efficaces et donc utiles pour mettre en réserve l'énergie des périodes d'abondance. Ces gènes ont été progressivement sélectionnés par nécessité car ils permettaient à ceux qui les possédaient de mieux survivre lorsque la disette survenait. En 2010, la nourriture est abondante (au moins dans les pays les plus développés), mais les gènes de stockage sont encore là, il leur faudrait des milliers d'années avant de disparaître, en admettant par ailleurs que l'abondance dure aussi longtemps ! Cette hypothèse dite du gène économe connaît de nombreuses exceptions et ne rend compte que d'une partie du phénomène, plusieurs des gènes identifiés comme étant liés à l'affection n'ayant en effet aucune relation avec le stockage des graisses, et en particulier des gènes exprimés dans

le système nerveux central, dont l'hypothalamus. L'obésité n'est pas une affection simple, elle est également liée à l'inflammation⁷ (NATHAN 2008), au microbiome, c'est-à-dire au contenu de la flore intestinale (LEVY 2005) et à bien d'autres facteurs. Alors, déterminisme darwinien de nature inconnue ou simple hasard, pas si simple.

C'est ici qu'intervient une notion très évolutionniste, mais qui a aussi une grande portée en médecine, la notion de convergence. En biologie de l'évolution, on appelle convergence le fait que des caractères semblables, génétiques ou morphologiques, puissent évoluer dans des espèces très éloignées sur un plan génétique (BARTON *et al.* 2007). Exemple : les oiseaux sont issus de la lignée des dinosaures, les chauves-souris sont des mammifères de la même lignée que les rats et les souris, les abeilles sont des insectes. Dinosaures, mammifères et insectes ont des origines distinctes, mais tous, oiseaux, chauve-souris et insectes, volent, tous ont en commun une morphologie acquise par des moyens distincts qui leur permettent de voler. La convergence signifie clairement convergence vers une même fonction.

La notion de convergence en matière d'évolution sous-entend bien évidemment une notion de temps qui n'a rien à voir avec l'échelle de temps que nous expérimentons tous les jours, mais en médecine humaine convergence signifie réparation, compensation, remplacement d'une fonction physiologique par un artifice qui va de la prothèse à la stimulation de la neuroplasticité. La confection d'une prothèse n'est jamais qu'une fonction humaine parmi d'autres. On parle aussi en génétique humaine de convergence (le chapitre est complexe, cf. HARRAP & Petrou 2001) lorsque la même fonction est altérée par des mutations survenues sur des gènes différents (dans certaines cardiomyopathies par exemple, la même infirmité peut être occasionnée par des mutations sur la myosine ou sur la protéine C, voire sur des canaux ioniques). Les exemples sont légion. Encore faut-il définir de quelle fonction précise il s'agit, c'est souvent le problème, la communication en est un bon exemple, communication inclut langage, mais pas seulement, on communique aussi par signes, par

7. On a en effet récemment démontré, entre autres, la présence en excès de macrophages, qui ont la même origine que les adipocytes, dans le tissu adipeux des obèses. Ceux-ci ont de ce fait des concentrations excessives de marqueurs plasmatiques de l'inflammation. C'est une des explications du risque cardiovasculaire des obèses (NATHAN 2008, SWYNGHEDAUW 2009).

Internet... Il appartient au médecin devant un défaut dans la communication, une aphasie par exemple, d'imaginer un autre moyen de communiquer, qui est l'essentiel, que la simple parole, et il y en a beaucoup.

2 ⇨ Y a-t-il une médecine évolutionniste (ou évolutionnaire)?

La médecine évolutionniste ou médecine darwinienne⁸ se définit comme l'étude du conflit entre notre génome, relativement stable et façonné par des milliards d'années d'environnement changeant, au cours duquel les périodes de disette ou de chaleur alternaient avec les périodes fastes ou froides, et l'environnement actuel qui est brutalement devenu plus chaud, pauvre en infections, riche en disponibilités alimentaires et en sel (NESSE & WILLIAMS 1994, TREVATHAN *et al.* 2007, STEARNS & KOELLA 2007, GLUCKMAN *et al.* 2009).

8. Le terme de médecine darwinienne couramment employé par les pionniers – anglo-saxons – de la discipline est difficile à accepter (voir aussi MÉTHOT 2009). Il présuppose en effet une vue déterministe uniquement darwinienne de l'évolution basée sur la seule pression sélective et l'adaptation, ce qui est maintenant difficile à défendre pour les raisons suivantes. (1) Au cours de l'évolution, l'adaptation biologique est toujours imparfaite ou incomplète (LECOINTRE 2009). (2) L'évolution neutre non adaptative, due au seul hasard et à la simple dérive génétique, représente une partie importante du processus évolutif dont l'évaluation est en cours (LYNCH 2007). (3) Une analyse détaillée des gènes dits de spéciation, à l'origine de plusieurs espèces de levures, de mouches, de souris et de plantes a montré, de façon surprenante, que les premières étapes de la spéciation étaient rarement adaptatives (PRESGRAVES 2010). (4) Plusieurs données fournies par l'évolution expérimentale (sur des colonies bactériennes) ont mis en évidence la complexité du couplage existant entre l'évolution des mutations et celle de l'adaptation (BARRICK *et al.* 2009). (5) Il y a des maladies génétiques qui confèrent à ceux qui en souffrent un certain avantage, mais seulement dans un certain environnement (les pays d'endémie paludéenne pour la drépanocytose). (6) La complexité des interférences gènes/environnement et gènes/gènes rend caduque toute tentative de simplification radicale, il faut par exemple tenir compte des symbioses que nous établissons avec les bactéries, parasites et virus qui nous infectent et dont les génomes évoluent et peuvent s'incorporer au génome humain ou en compléter l'activité sur un plan métabolique (GILL *et al.* 2006, GERSTEIN *et al.* 2007). (7) On connaît maintenant des mécanismes, en partie transmissibles, sans altérations de la structure du génome, les mécanismes épigénétiques. Ils jouent un rôle important dans le développement mais aussi dans nombre d'affections (CUZIN *et al.* 2008).

Si le génotype n'est finalement que l'outil que chacun possède et qui lui permet de répondre à l'environnement et à ses variations, la maladie est un phénotype particulier résultant d'un conflit entre notre génome et son environnement. Le concept de norme de réaction génotypique, cher aux biologistes de l'évolution et aux généticiens (DEWITT & SCHEINER 2004, SWYNGHEDAUW 2006), est un outil que l'on peut adapter à l'étude de la physiopathologie. Pour les biologistes de l'évolution, cet outil permet de quantifier la plasticité d'un phénotype face à des changements d'environnement, comme la température externe, et d'analyser ces effets sur différents génotypes. En médecine, le phénotype, c'est le trait, le signe clinique, lequel dépend toujours des interactions entre génotype et environnement dont il n'est que l'expression, et la maladie apparaît pour un certain seuil défini à la fois par le génotype et par l'environnement (SWYNGHEDAUW 2009). Il existe de véritables « maladies de l'adaptation » (LOMPRE 1979, SWYNGHEDAUW 1999).

2.1 ➔ Conséquences médicales du réchauffement climatique

2.1.1 ➔ Données cliniques et épidémiologiques

Le réchauffement climatique et son origine anthropogénique ne font plus guère de doute (Éditorial 2007). Ses conséquences médicales en termes épidémiologiques sont assez bien connues (le détail des références se trouve dans SWYNGHEDAUW 2008 et BESANCENOT 2007). (1) La simple élévation de la température moyenne du globe s'accompagne, en règle générale, d'une augmentation de la mortalité et de la morbidité globale. La courbe mortalité/température externe est une courbe en J, mais la pente de la branche chaude du J est plus prononcée que celle de la branche froide, ces données étant très dépendantes du niveau de vie. Une étude récente portant sur 50 villes a confirmé que la mortalité globale, cardiovasculaire en particulier, augmente aux températures extrêmes. (2) Les effets des vagues de chaleur comme celle qui a touché la France en août 2003 et d'autres pays, ont fait l'objet de plusieurs études détaillées. Le fait que la mortalité due aux grosses vagues de chaleur d'août 2003 n'ait pas (ou peu) été suivie d'une diminution de la mortalité (*harvesting effect*) suggère qu'il s'est agi d'un effet thermique direct – le coup de chaleur – indépendant du niveau de santé antérieur mais fortement lié à l'âge, et que, en d'autres termes, une telle mortalité reflète surtout les limites de nos capacités d'adaptation aux changements de la température externe. (3) Les variations climatiques ont participé aux

modifications dans la répartition (surtout en altitude) et la virulence de certains agents pathogènes parasites (dengue, paludisme, etc.) et de leurs vecteurs, bien que cet effet soit en grande partie compensé par les mesures de protection instaurées par l'homme (GETHING *et al.* 2010). (4) Les effets indirects dus aux variations extrêmes du cycle de l'eau, aux changements dans la fréquence et l'intensité des cyclones tropicaux, aux changements de la biodiversité et de l'écosystème ont eu et auront des conséquences en termes de niveau de vie, lequel est un facteur de risque médical considérable.

2.1.2 ➔ Mécanismes

Ces données posent un certain nombre de questions purement biologiques rarement soulevées. Le métabolisme des êtres vivants a en effet été façonné par un environnement relativement froid et les limites précises de son adaptation à un changement thermique, brutal à l'échelle de l'évolution, sont inconnues. La régulation de la température interne se fait dans l'immédiat au moyen de la sudation et de la vasodilatation cutanée, mais la régulation à long terme dépend essentiellement des nombreux mécanismes de découplage de l'oxydation phosphorylante mitochondriale. À cela s'ajoutent les éléments responsables du coup de chaleur et en particulier la libération de cytokines pro-inflammatoires, récemment mise en évidence. Le degré d'adaptabilité de ces mécanismes à une élévation thermique est pour le moment totalement inconnu et aucun travail d'envergure n'est vraiment entrepris pour le résoudre.

2.2 ➔ L'épidémie de maladies auto-immunes et allergiques

En réponse à une blessure, une infection ou un stress, les tissus mettent en œuvre toute une batterie de signaux dont la finalité est la réparation du dommage causé, la protection contre la possible invasion bactérienne et le rétablissement de l'homéostasie perturbée. L'inflammation a été sélectionnée au cours de l'évolution parce qu'elle est d'abord adaptative, mais la réaction peut avoir des conséquences délétères surtout lorsqu'elle passe à la chronicité dans un environnement différent de celui pour lequel elle a été sélectionnée (MEDZITOV 2008, BACH 2002).

2.2.1 ➔ Données environnementales et épidémiologiques

L'incidence de maladies allergiques comme l'asthme, le rhume des foins ou la dermatite atopique, et celle des maladies auto-immunes comme

le diabète type 1⁹, l'iléite de Crohn et le cadre très général des maladies inflammatoires de l'intestin, la sclérose en plaques, a pratiquement doublé ou triplé depuis 1950 (BACH 2002). Géographiquement, l'incidence des maladies allergiques et auto-immunes diminue du Nord vers le Sud. Ce gradient est attribué à la conjonction statut économique faible + températures extérieures élevées. L'influence de l'environnement y est donc majeure. Le diabète de type 1 est par exemple une rareté au Pakistan (1 pour 100 000), mais il est devenu aussi fréquent chez les immigrants pakistanais en Grande-Bretagne que chez les Britanniques d'origine (autour de 10-11 pour 100 000). L'incidence du diabète de type 1, comme celle de la sclérose en plaques et de l'asthme, est très significativement corrélée au produit intérieur brut, c'est-à-dire à la qualité des soins et à la prévention des infections. La protection apportée par des infections extérieures, par exemple helminthiques, est également bien documentée.

L'hypothèse dite hygiéniste repose sur la constatation que l'augmentation récente et massive de l'incidence de ce type d'affections dans les pays développés est associée à la réduction des infections bactériennes, virales et/ou parasitaires. Elle a fait l'objet d'une revue très complète récente, cette relation est négative et réversible. L'antiparallélisme est à la fois chronologique et géographique et plusieurs mécanismes biologiques, crédibles, ont été proposés pour l'expliquer (WILLS-KARP *et al.* 2001, BACH 2002).

2.2.2 ➔ Données génétiques

Tout le monde ne fait pas d'asthme ou de maladie de Crohn ou de diabète de type 1, il faut pour cela posséder un certain nombre de facteurs génétiques que l'on commence à connaître, même si le bilan en est encore incomplet (THE WELLCOME TRUST 2007) et si le mécanisme d'action n'en est encore que partiellement connu.

Les maladies inflammatoires de l'intestin (dont la colite ulcéreuse et l'iléite de Crohn) sont une cause extrêmement fréquente de consultation en médecine générale; elles ont toutes une composante génétique

9. Il y a deux types de diabète : un type rare, le diabète de type 1, survenant surtout chez les sujets jeunes, grave, insulino-dépendant, et qui est clairement une maladie auto-immune, et le diabète de type 2 qui est de loin le plus fréquent et survient beaucoup plus tard; il est également appelé diabète gras car étroitement lié à la surcharge pondérale.

qui semble bien être responsable de la spécificité de l'affection (GAYA *et al.* 2006). Les progrès effectués en matière d'hygiène favoriseraient ce type d'affection en rendant la barrière intestinale moins préparée aux agressions et aux entrées bactériennes, mais l'association avec un ou plusieurs facteurs génétiques paraît indispensable, et leur recensement est actuellement en cours et a déjà mis en évidence plusieurs associations fortes avec des gènes codant plusieurs partenaires de la cascade inflammatoire.

Les gènes de susceptibilité liés au diabète de type 1 sont surtout les gènes d'une des branches du système immunitaire, plusieurs études récentes ont permis de préciser les gènes de susceptibilité nécessaires et suffisants pour que ce type de diabète apparaisse (BLUESTONE *et al.* 2010). Le même type de résultats a été observé avec l'asthme; le journal *The Lancet* a consacré un numéro spécial entier à cette nouvelle manière de considérer cette affection (WENZEL 2006).

2.2.3 ➤ Mécanismes

Le pourquoi de cette relation négative se doit de tenir compte de plusieurs facteurs comme la concentration et la durée d'exposition, mais surtout il faut également articuler le facteur environnemental à la composante génétique dont on ne connaît encore que partiellement les déterminants.

L'analyse de la flore intestinale (appelée microbiome) et de son génome («notre autre génome», page de couverture du numéro de *Nature* du 4 mars 2010) est en plein développement (QIN *et al.* 2010). Le tube digestif et son épithélium sont un lieu privilégié pour une telle confrontation et de multiples données expérimentales sont en faveur de l'hypothèse selon laquelle une hygiène de vie excessive modifie radicalement le système immunitaire qui nous protège normalement contre notre flore intestinale. L'hypothèse hygiéniste a d'abord été proposée pour expliquer l'incidence des maladies allergiques, mais récemment il est devenu évident que les maladies auto-immunes et allergiques étaient jumelles et souvent présentes chez un même patient et que leur origine était commune à la fois sur un plan épidémiologique et mécanistique.

3 ➤ Évolution et politiques de santé

Les exemples précédents ne sont pas seuls, bien d'autres peuvent être donnés concernant le cancer, forme accélérée de l'évolution ou les

maladies cardiovasculaires (SWYNGHEDAUW 2009). Réconcilier la médecine avec l'évolution peut avoir un impact qui peut surprendre, cet impact concerne la politique de santé et les prévisions qu'elle se doit de faire. Les exemples qui ont été donnés concernent en effet des données impliquant la santé publique et l'on peut prédire sans peine qu'ils seront à la une de nos préoccupations futures. Sauf changement radical dans les politiques mondiales concernant les émissions de gaz à effet de serre, on peut en effet s'attendre à une aggravation des changements climatiques dans les années qui vont suivre. Ainsi, la biologie de thermogénèse et la physiopathologie du choc thermique deviendront rapidement des priorités, ce qu'ils ne sont pas actuellement. L'évolution des résistances bactériennes, parasitaires et virales aux agents anti-infectieux, celle de l'efficacité des vaccins et de l'ensemble d'une politique de l'hygiène se doivent d'être considérés en tenant compte de la dysrégulation du système immunitaire qu'elles ne peuvent manquer d'induire. La prévention de l'obésité et la promotion de l'activité physique régulière sont déjà des objectifs de santé publique, par contre les facteurs de risque génétiques qui leur sont liés ne sont pas encore devenus les indicateurs qu'ils pourraient être. La biologie de l'évolution pourrait constituer le socle scientifique d'une politique de santé plus rationnelle.

4 ↻ Conclusion et résumé

Le but de cet article est d'essayer de réconcilier médecine et évolution biologique, et de contribuer à la promotion de ce type d'enseignement. Par médecine évolutionniste, on entend tout ce qui résulte du conflit entre notre génome tel qu'il a été façonné par un environnement souvent hostile durant des millions d'années et l'environnement actuel qui a été modifié par l'activité humaine dans ce qu'elle a à la fois de bénéfique mais aussi de délétère. L'évolution biologique est un processus irréversible, en cours, discontinu, caractérisé par des périodes de stase, suivies de ponctuations. Elle a un déterminant, les mutations géniques, et procède par pression sélective darwinienne et au hasard des dérives génétiques.

Le fait médical peut être considéré comme étant le résultat d'un conflit génome/environnement. Il peut n'être que génétique dans certaines maladies monogéniques, ou n'être qu'environnemental, après un accident grave par exemple. Néanmoins dans la très grande majorité des maladies dites communes, le tableau clinique est défini par un

conflit entre ces deux critères, le critère génétique n'étant généralement qu'incomplètement connu. Quelques exemples en ont été choisis. (1) Les conséquences médicales de l'effet de serre. L'absence de surmortalité après les récentes vagues de chaleur suggère que l'élément déterminant en a été le coup de chaleur et l'âge. Les effets au long cours de l'élévation globale de la température mettent au premier plan la physiologie de la thermolyse, chapitre mal connu de la physiologie. (2) L'hypothèse hygiéniste postule l'existence d'une relation inverse entre la diminution des infections et l'accroissement exponentiel de l'incidence des maladies auto-immune et allergiques. Le substratum génétique de ce type d'affections est actuellement en cours de décryptage. (3) L'épidémie d'obésité et de diabète qui se développe en ce moment dans nos pays doit être considérée comme résultant d'un conflit entre un environnement trop riche en calories et des variants des gènes régulant l'appétit.

Considérer le fait médical à la lueur de l'évolution peut être une nouvelle manière d'établir les prémices d'une politique de santé globalisée.

☉Références bibliographiques☉

A

AMEISEN J.-C. (2002), "On the origin, evolution, and nature of programmed cell death: a timeline of four billion years", *Cell Death Differ*, 9: 367-393.

B

BACH J.-F. (2002), "The effect of infections on susceptibility to autoimmune and allergic diseases", *N. Engl. J. Med.*, 347: 911-920.

BARRICK J.E., YU D.S., YOON S.H. *et al.* (2009), "Genome evolution and adaptation in a long-term experiment with *Escherichia coli*", *Nature*, 461: 1243-1247.

BARTON N.H., BRIGGS D.E.G., EISNE J.A., GOLDSTEIN D.B. & PATEL N.H. (2007), *Evolution*, New York, Cold Spring Harbor Laboratory Press.

BESANCENOT J.-P. (2007), *Notre santé à l'épreuve du changement climatique*, Paris, Delachaux et Niestlé.

BLUESTONE J.A., HEROLD K. & EISENBARTH G. (2010), "Genetics, pathogenesis and clinical interventions in type 1 diabetes", *Nature*, 464: 1293-1300.

C

CUZIN F., GRANDJEAN V. & RASSOULZADEGAN M. (2008), "Inherited variation at the epigenetic level: paramutation from the plant to the mouse", *Curr. Opin. Genet. Dev.*, 18: 193-196.

D

DAVIDSON C.J., HIRT R.P., KALPANA L.A.L. *et al.* (2003), "Molecular evolution of the vertebrate blood coagulation network", *Thromb Haemostasis*, 89: 420-428.

DEWITT T.J. & SCHEINER S.M. (2004), *Phenotypic plasticity. Functional and conceptual approaches*, New York, Oxford University Press.

DONNELLY P. (2008), "Progress and challenges in genome-wide association studies in humans", *Nature*, 456: 728-731.

E

Éditorial (2007), "Light at the end of the tunnel", *Nature*, 445: 567.

F

FIDOCK D.A. (2010), "Priming the antimalarial pipeline", *Nature*, 465: 297-298.

G

GARG S., NARULA J. & CHANDRASHEKHAR Y. (2004), "Apoptosis and heart failure: clinical relevance and therapeutic target", *J. Mol. Cell Cardiol.*, 38: 73-79.

GAYA D.R., RUSSELL R.K., NIMMA E.R. *et al.* (2006), "New genes in inflammatory bowel disease: lessons for complex diseases?", *Lancet*, 367: 1271-1284.

GERSTEIN M.B., BRUCE C., ROZOWSKY J.S. *et al.* (2007), "What is a gene, post-ENCODE? History and updated definition", *Genome Res.*, June 18: 669-681.

GETHING P.W., SMITH D.L., PATIL A.P. *et al.* (2010), "Climate change and the global malaria recession", *Nature*, 465: 342-345.

GILL S.R. *et al.* (2006), "Metagenomic analysis of the human distal gut microbiome", *Science*, 312: 1355-1359.

GLUCKMAN P., BEEDLE A. & HANSON M. (2009), *Principles of evolutionary medicine*, Oxford, Oxford University Press.

H

HANAHAH D. & WEINBERG R.A. (2000), "The hallmarks of cancer", *Cell*, 100: 57-70.

HARRAP S.B. & PETROU S. (2001), "Utility of genetic approaches to common cardiovascular diseases", *Am. J. Physiol. Heart. Circ. Physiol.*, 281: H1-H6.

J

JANCOVICI J.-M. & GRANDJEAN A. (2009), *C'est maintenant! Trois ans pour sauver le monde*, Paris, Seuil.

JOBLING M.A., HURLES M.E. & TYLER-SMITH C. (2004), *Human evolutionary genetics. Origins, peoples and disease*, New York, Abingdon, Garland Science.

JORDAN B. (2010), «À la recherche de l'héritabilité perdue», *Médecine/Sciences*, 26: 541-543.

JOUZEL J. & DEBROISE A. (2007), *Le climat: jeu dangereux. Dernières nouvelles de la planète*, Paris, Dunod.

K

KAPLAN J.-C. & DELPECH M. (2007), *Biologie moléculaire et médecine*, Paris, Médecine-Sciences Flammarion.

L

LECOINTRE G. (dir.) (2009), *Guide critique de l'évolution*, Paris, Belin.

LECOINTRE G. & LE GUYADER H. (2001), *Classification phylogénétique du vivant*, Paris, Belin.

LEY R.E., BÄCKHED F., TURNBAUGH P. *et al.* (2005), "Obesity alters gut microbial ecology", *Proc Natl Acad Sci USA*, 102: 11070-11075.

LOMPÉRÉ A.M., SCHWARTZ K., D'ALBIS A., LACOMBE G., THIEM N.Y. & SWYNGHEDAUF B. (1979), "Myosin isoenzyme redistribution in chronic heart overloading", *Nature*, 282, 105-107.

LYNCH M. (2007), "The frailty of adaptive hypotheses for the origins of organismal complexity", *Proc. Natl. Acad. Sci. USA*, 104: 8597-8604.

M

MCPHERSON R. (2010), "Chromosome 9p21 and coronary artery disease", *N. Engl. J. Med.*, 362: 1736-1737.

MEDZHITOV R. (2008), "Origin and physiological roles of inflammation", *Nature*, 454: 428-435.

MÉTHOT P.-O. (2009), «Darwin et la médecine: intérêt et limites des explications évolutives en médecine», in T. Heams, P. Huneman, G. Lecointre & M. Silberstein (dir.), *Les Mondes darwiniens*, Paris, Syllepse.

MONOD J. (1970), *Le Hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*, Paris, Seuil.

N

NATHAN C. (2008), "Epidemic inflammation: pondering obesity", *Mol. Med.*, 14: 485-492.

NESSE R.M. & WILLIAMS G. (1994), *Why we get sick: the new science of Darwinian medicine*, New York, Times Books.

NESSE R.M., BERGSTROM C.T., ELLISON P.T. *et al.* (2009), "Making evolutionary biology a basic science for medicine", *Proc. Natl Acad. Sc. USA*, 107 (suppl 1): 1800-1807.

NÜMBERGER T., BRUNNER F., KEMMERLING B. *et al.* (2004), "Innate immunity in plants and animals: striking similarities and obvious differences", *Immunological Rev.*, 198: 249-266.

P

PRESGRAVES D.C. (2010), "The molecular evolutionary basis of species formation", *Nature Rev. Genetics*, 11: 175-180.

Q

QIN J., LI R., ARUMUGAM M. *et al.* (2010), "A human gut microbial gene catalogue established by metagenomic sequencing", *Nature*, 464: 59-65.

R

READ A.F. & HUIJGEN S. (2009), "Evolutionary biology and the avoidance of antimicrobial resistance", *Evolutionary Applications*, 2: 40-51.

S

SMALL E.M., FROST R.J.A. & OLSON E.N. (2010), "MicroRNAs add a new dimension to cardiovascular disease", *Circulation*, 121: 1022-1032.

STEARNS S.C. & KOELLA J.C. (2007), *Evolution in health and disease*, Oxford University Press.

STRATTON M.R., CAMPBELL P.J. & FUTREAL A. (2009), "The cancer genome", *Nature*, 458: 719-724.

SWYNGHEDAUB B. (1986), "Developmental and functional adaptation of contractile proteins in cardiac and skeletal muscle", *Physiol. Rev.*, 66: 710-771.

SWYNGHEDAUB B. (1999), "Molecular mechanisms of myocardial remodeling", *Physiol. Rev.*, 79: 215-262.

SWYNGHEDAUB B. (2006), "Phenotypic plasticity of adult myocardium. Molecular mechanisms", *J. Exp. Biol.*, 209: 2320-2327.

SWYNGHEDAUB B. (2008), «Conséquences médicales de l'effet de serre», *La Presse Médicale*, 38: 551-561.

SWYNGHEDAUB B. (2009), *Quand le gène est en conflit avec son environnement. Introduction à la médecine darwinienne*, Bruxelles, De Boeck.

T

THE WELLCOME TRUST CASE CONTROL CONSORTIUM (2007), "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls", *Nature*, 447: 661-678.

TREVATHAN W.R., SMITH E.O. & MCKENNA J.J. (2007), *Evolutionary medicine. New perspectives*, Oxford University Press.

W

WENZEL S.E. (2006), "Asthma: defining of the persistent adult phenotypes", *Lancet*, 368: 804-813.

WILLS-KARP M., SANTELIZ J. & KARP C.L. (2001), "The germless theory of allergic disease: revisiting the hygiene hypothesis", *Nature Rev. Immunol.*, 1: 69-75.

Y

YAZDANBAKHSH M., KREMANER P.G. & VAN REE R. (2002), "Allergy, parasites, and the hygiene hypothesis", *Science*, 296: 490-494.

Z

ZIMMER C. (2010), *The tangled bank*, Roberts and company.

☞ Notices biographiques ☞

Bernard Swynghedauw. Docteur en médecine, docteur-ès-sciences, ancien interne des hôpitaux de Paris, directeur de recherches à l'Inserm (émérite), past-president of the Federation of the European Physiological Societies (FEPS), membre correspondant de l'Académie nationale de médecine. U942-INSERM, Hôpital Lariboisière, Paris.

Bernard.Swynghedauw@inserm.fr

Christian Frelin. Docteur-ès-sciences, directeur de recherches au CNRS, UMR 6543-Université de Nice Sophia-Antipolis.

cfrelin@unice.fr

Avertissement. Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur et des auteurs. Les textes et illustrations figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage personnel, scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document selon ce modèle :

Bernard Swynghedauw et Christian Frelin, « L'évolution biologique en médecine. Un outil de formation, un moyen de hiérarchiser les informations et une base rationnelle à une politique de santé », in Gérard Lambert et Marc Silberstein (dir.), Matière première. Revue d'épistémologie [en ligne]. Nouvelle série, N° 1/2010 : Épistémologie de la médecine et de la santé, Éditions Matériologiques. Mis en ligne le 7 octobre 2010. URL : www.materiologiques.com

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.



Joël Coste **Inférences causales et probabilistes dans le raisonnement diagnostique : un éclairage historique sur le débat contemporain**

Depuis maintenant plus de trente ans, on lit régulièrement dans la presse médicale que seules les approches probabilistes *pourraient* gérer l'augmentation « explosive » du savoir diagnostique¹. Quoiqu'il en soit de ce savoir – aujourd'hui alimenté à flux continu par les nouveaux « examens complémentaires », produits souvent éphémères d'une certaine médecine technologique et commerciale –, l'association presque invariable du calcul probabiliste au verbe « pouvoir » et à l'idée de *potentialité*

1. « Nous déclarons qu'appliquer une technique probabiliste peut améliorer considérablement la précision du diagnostic anatomo-pathologique et peut grandement faciliter la communication entre anatomo-pathologistes et cliniciens. L'analyse probabiliste peut être aussi d'une utilité considérable pour l'interprétation et la présentation des résultats d'examens radiologiques ou isotopiques. » (SCHWARTZ, WOLFE & PAUKER 1981.)

« Les connaissances médicales augmentent de manière explosive. Alors que la disponibilité des informations pertinentes peut rendre l'exercice du diagnostic plus exact, celles-ci deviennent aussi de plus en plus difficiles à maîtriser par le médecin [...]. Les réseaux bayésiens peuvent apporter une aide à la décision en radiologie car ils peuvent modéliser l'incertitude, calculer et expliquer les probabilités post-test et prendre en compte efficacement une grande quantité d'information. » (BURNSIDE 2003.)

suggère quelques *résistances* de la réalité, du *terrain* de la pratique médicale, à ce calcul. Prenant du champ vis-à-vis du débat – passionné et souvent passionnel, mais parfois aussi très technique – sur le raisonnement probabiliste qui agite la communauté médicale depuis plus de trois décennies², nous voudrions dans cet article analyser une des raisons des *résistances* à sa mise en œuvre, à savoir l'existence de raisonnements alternatifs, depuis longtemps adaptés à la réalité de la pratique diagnostique: le raisonnement causal et le raisonnement physiopathologique, c'est-à-dire le raisonnement sur les *causes* ou sur les *mécanismes* des maladies. Après avoir brièvement présenté les avatars contemporains de ces trois raisonnements, et examiné leur efficacité et leurs limites dans leur pratique actuelle, nous solliciterons l'histoire de la médecine qui nous rappellera que le raisonnement diagnostique et le raisonnement causal furent étroitement intriqués pendant toute l'histoire de la médecine rationnelle – une intrication que l'on verra consubstantielle, et même fondatrice, de celle-ci.

L'histoire de la médecine nous rappellera aussi qu'il exista, dès les origines de la médecine rationnelle, des tensions autour des raisonnements et approches «déterministes» de la pratique médicale, notamment dans la conduite de la démarche diagnostique, et que ces tensions contribuèrent probablement au développement d'approches qualifiées d'«empiriques», qui ne sont pas sans parenté avec l'approche probabiliste contemporaine. Nous n'envisagerons bien entendu que quelques *moments* de l'histoire de la médecine et quelques *aspects* d'une histoire du diagnostic médical qui reste encore largement *terra incognita* des historiens.

1 ↻ Les avatars contemporains du raisonnement probabiliste, du raisonnement causal et du raisonnement physiopathologique pour le diagnostic médical³

Le diagnostic est «l'art de reconnaître les maladies par leurs symptômes et de les distinguer les unes des autres» (Littré). La *reconnaissance* des maladies repose sur la comparaison des symptômes et signes obser-

2. Pour un exemple des débats précoces sur le raisonnement probabiliste, cf. RANSOHOFF & FEINSTEIN (1976).

3. Cette première section est construite en grande partie à partir des chapitres 4, 10 et 13 du livre écrit avec Jean-Baptiste Paolaggi (1928-2010), *Le Raisonnement médical. De la science à la pratique clinique* (2001).

vés chez un patient donné à un ensemble de symptômes et signes servant de référence et réalisant la *représentation conceptuelle* de l'entité pathologique à laquelle le médecin pense être confronté. *Pour pouvoir reconnaître, le médecin doit bien sûr connaître* et avoir à disposition dans sa mémoire les connaissances relatives à la nosologie des affections (pour la question très importante des constructions nosologiques, voir *infra*). Parfois, le diagnostic est d'une *extrême facilité*, de par la nature de l'affection en cause qui est simple, ou du fait de l'expérience du médecin qui est grande. Certains peuvent même être faits par des profanes comme, par exemple, un furoncle ou une fracture diaphysaire ouverte. Le plus souvent toutefois, le diagnostic se construit progressivement, sur la base des informations recueillies lors de l'observation structurée du patient, par la mise en œuvre de *raisonnements*, qui permettent d'aller jusqu'à un certain stade de la démarche où le praticien pense avoir obtenu un accord satisfaisant entre les signes observés et ceux d'une entité pathologique définie (il se peut aussi qu'un diagnostic satisfaisant ne puisse pas être d'emblée réalisable du fait de la pauvreté des symptômes présents à un certain stade de l'évolution de la maladie et/ou parce qu'il est nécessaire de disposer d'autres informations : il faudra alors que le médecin se satisfasse de prendre les mesures permettant de sauvegarder la vie ou la fonction en attendant de pouvoir progresser de façon plus satisfaisante).

1.2 ➔ Les principaux raisonnements utilisés dans la démarche diagnostique

Les raisonnements mis en œuvre par le médecin dans la démarche diagnostique sont divers (et parfois multiples pour un même cas). Un premier groupe de raisonnements, à peu près admis par l'ensemble de la communauté médicale scientifique, comprend le raisonnement descriptif, le raisonnement causal, le raisonnement physiopathologique et le raisonnement probabiliste. Un second groupe comprend des modes de raisonnement plus controversés : le raisonnement analogique et le raisonnement basé sur, ou plutôt aidé par des algorithmes ou des *règles de prédiction clinique*⁴.

4. À côté de ces raisonnements explicites et conscients, les médecins ont aussi fréquemment recours en pratique clinique à des *infra-raisonnements*, appelés «heuristiques», et principalement à l'heuristique de représentativité (très

Le *raisonnement probabiliste* recourt au calcul des probabilités pour déterminer la « probabilité post-test » d'une maladie, c'est-à-dire la probabilité de cette maladie après la réalisation d'un test diagnostique. Les probabilités $P(M+/T+)$, si le test est positif, et $P(M+/T-)$, si le test est négatif, peuvent être calculées, grâce au théorème de Bayes, à partir de la probabilité initiale ou « pré-test » de la maladie, $P(M+)$, et des caractéristiques intrinsèques du test : la sensibilité (Se) qui est aussi la probabilité que le test soit positif chez le sujet avec la maladie, et la spécificité (Sp) qui est la probabilité que le test soit négatif chez le sujet indemne de la maladie :

$$P(M+ / T-) = \frac{(1 - Se) \times P(M+)}{(1 - Se) \times P(M+) + Sp \times (1 - P(M+))}$$

$$P(M+ / T-) = \frac{(1 - Se) \times P(M+)}{(1 - Se) \times P(M+) + Sp \times (1 - P(M+))}$$

Malgré sa formulation mathématique et sa simplicité, le raisonnement probabiliste reste extrêmement peu utilisé *directement* en pratique médicale de tous les jours. Les difficultés matérielles de sa mise en œuvre (la nécessité du recours à l'ordinateur ou, au mieux, à la calculatrice) et les connaissances insuffisantes des caractéristiques métrologiques des tests diagnostiques, même les plus élémentaires, sont généralement invoquées pour expliquer cette très faible utilisation. Le raisonnement probabiliste est en revanche quelquefois utilisé *indirectement*, depuis une quinzaine d'années, par l'intermédiaire des *règles de prédiction clinique* dérivées de la modélisation de situations cliniques avec des modèles probabilistes ; ces règles permettent de retenir ou d'éliminer une hypothèse diagnostique en fonction de la réponse à une petite série de critères ou de la valeur d'un score construit avec ces critères⁵. Par exemple, la règle dite d'Ottawa permet d'éliminer

proche de la reconnaissance de forme), à l'heuristique de disponibilité et à l'heuristique d'ajustement (qui sont en fait des modes de gestion implicite des probabilités). Sur les heuristiques, qui dépassent le cadre de cet article, cf. notamment TVERSKY & KAHNEMAN (1974).

5. Ces règles de prédiction clinique ont été particulièrement encouragées par les promoteurs de l'*Evidence Based Medicine*. Sur cette approche de l'apprentissage et de la pratique de la médecine, présentée en 1992, cf. COSTE & PAOLAGGI (2001, p. 141-145).

une fracture de la cheville à la suite d'un traumatisme si le sujet peut tenir un appui monopodal (du côté traumatisé) et si la palpation des deux régions malléolaires ne réveille pas de douleurs intenses (STIELL *et al.* 1994).

Le *raisonnement causal* est principalement mis en œuvre pour faire un diagnostic étiologique, c'est-à-dire attribuer un syndrome observé à une cause parmi celles identifiées par la science médicale, cause qui peut être infectieuse, tumorale, vasculaire, traumatique, inflammatoire, etc., et pour décider des « indications » de traitements à visée curative, cet usage n'étant d'ailleurs que la suite logique de la démarche précédente (l'action sur la cause entraînant logiquement une suppression ou une réduction de l'effet pathologique)⁶. Le raisonnement causal *sensu stricto* considère la cause des phénomènes morbides sans prendre en compte le mécanisme intime des troubles et reste donc en amont des connaissances physiopathologiques, ce qui peut être parfois fâcheux. Par exemple, un germe peut être la cause d'une infection bien que les troubles entraînés par la pénétration et le développement du germe dans l'organisme varient considérablement d'un sujet à l'autre, d'un viscère à l'autre, d'un état immunologique à un autre, et ici ne pas prendre en compte ces mécanismes physiopathologiques serait dommageable (un bon exemple est la toxoplasmose, dont les manifestations varient, chez l'enfant et l'adulte, la femme enceinte, le sujet immunodéprimé). De plus, le raisonnement causal est évidemment inopérant pour les (nombreuses) affections, incontestables par leur tableau clinique, caractérisées par des lésions bien définies et dont les événements physiopathologiques sont en partie éclaircis, mais dont la cause exacte n'est pas élucidée...

Le *raisonnement physiopathologique* comprend la mise en œuvre de connaissances issues des disciplines cliniques et fondamentales *combinaées* pour la compréhension des mécanismes des troubles pathologiques. Ces disciplines fondamentales dont on utilise les connaissances peuvent être – et ont été, successivement et cumulativement – la physiologie des fonctions, la médecine expérimentale, la bactériologie, la physiologie cellulaire, la biophysique, la génétique, la biologie moléculaire, etc. Dans le domaine du diagnostic, le raisonnement physiopathologique est très largement employé et effectivement utile dans de nombreuses

6. Sur l'épistémologie de la problématique de l'étiologie en médecine, cf. les articles de Steeves Demazeux et d'Élodie Giroux, dans ce volume. (NdÉ.)

situations. En neurologie, le siège des lésions du système nerveux peut être déduit de l'observation des troubles et de leur mise en relation avec les fonctions des aires ou des noyaux nerveux et le trajet des différentes voies nerveuses. Par exemple, des troubles de la marche associés à un déficit moteur d'un membre inférieur sur lequel on trouve des réflexes vifs et diffusés, un signe de Babinski (qui indiquent un « syndrome pyramidal », du nom des voies nerveuses motrices), à un déficit de la sensibilité douloureuse de l'autre membre et à quelques troubles sphinctériens orientent immédiatement vers une atteinte de moelle épinière dorsale ou lombaire. De même, le diagnostic d'une anémie peut être conduit, une fois celle-ci reconnue, par un raisonnement sur ses différents mécanismes, en prenant en compte la taille des globules rouges, le taux de fer sérique et de ferritine, ainsi que des signes associés d'une hémolyse; en endocrinologie, la connaissance des rétrocontrôles permet d'orienter immédiatement vers des origines centrales, hypothalamo-hypophysaires, ou périphériques pour les hyper/hypothyroïdies, hyper/hypocorticismes, etc. Le raisonnement physiopathologique a aussi permis de donner la clé pour le diagnostic étiologique de maladies rares comme les rachitismes ou les ostéomalacies dites vitamino-résistantes. De même, le raisonnement physiopathologique (mis en œuvre par GARROD 1859 et par BALL 1971) permet par exemple d'incriminer le rôle de l'accumulation du plomb et le saturnisme chronique dans les « épidémies » de goutte touchant les buveurs de porto anglais au XIX^e siècle (porto fortifié au plomb pour le voyage) ou les distillateurs clandestins de whiskey aux Etats-Unis dans les années 1960: l'hyperuricémie importante à l'origine des manifestations de goutte fut en effet expliquée d'une part par une atteinte rénale (la néphropathie tubulo-interstitielle toxique) entraînant une diminution de l'excrétion urinaire d'acide urique, et d'autre part par une perturbation du métabolisme des purines par l'effet toxique direct du plomb, responsable d'une élévation secondaire des taux d'acide urique. À côté de brillants succès de ce type, le raisonnement physiopathologique a toutefois été souvent trahi par la confusion entre hypothèses et faits démontrés en biologie. Il est en effet parfois difficile, dans le cadre de la recherche médicale biologique, d'établir de façon rigoureuse la preuve expérimentale d'une hypothèse physiopathologique, et les biologistes ont quelquefois fait preuve d'un enthousiasme exagéré pour leurs découvertes supposées...

fragilisant ainsi les constructions nosologiques sur lesquelles raisonnent les médecins.

1.3 ↻ Les entités et les constructions nosologiques contemporaines

Dans *Le Raisonnement médical, de la science à la pratique clinique*, nous avons proposé de distinguer plusieurs groupes d'entités ou de constructions nosologiques (tableau 1) suivant le niveau de validation de leurs éléments de définition, particulièrement de leur cause et mécanisme physiopathologique.

Le premier groupe comprend les maladies totalement définies, par exemple une affection monogénique dont on a pu préciser toute la séquence des événements depuis l'ADN jusqu'aux symptômes et à leurs conséquences sur l'évolution (la mucoviscidose en est un bon exemple⁷).

Le deuxième groupe est essentiellement constitué des maladies infectieuses aux agents identifiés, bactériens (la typhoïde, par exemple), viraux (le VIH, par exemple) ou parasitaires (le paludisme, par exemple). Certains cancers aux causes virales établies (le cancer de l'utérus par exemple, causé par certains types de papillomavirus) appartiennent aussi à ce groupe. Pour ces affections, un mécanisme causal et une physiopathologie précise ont été établis.

Les autres affections tumorales (où la prolifération cellulaire constitue le principe même de la maladie) forment le troisième groupe, avec certaines affections hématologiques, endocrinologiques ou métaboliques et la nécrose tissulaire ischémique. Dans ce groupe, les mécanismes physiopathologiques sont bien identifiés mais les causes à l'origine du déclenchement des troubles ne le sont pas.

Le quatrième groupe est formé des maladies connues par leurs symptômes, leur évolution, par certaines de leurs lésions, mais dont la nature profonde et les raisons intimes de leur déclenchement restent inconnues. C'est le cas, par exemple, de maladies comme la sclérose en plaques ou la polyarthrite rhumatoïde, dont personne ne mettrait en doute l'existence et la solidité de la spécification mais dont la cause et la nature intime même des mécanismes ne font, pour l'instant, que l'objet de supputations ou de démonstrations très imparfaites.

7. Sur cette pathologie et les recherches en génétique qui y sont associées, cf. l'article de Catherine Dekeuwer, dans ce volume. (NdÉ.)

Les deux derniers groupes sont constitués par les maladies définies uniquement par leurs symptômes. En effet, il peut arriver qu'une affection ne puisse se définir que par l'association de symptômes, ou «syndrome», et que les lésions ne soient pas caractérisées avec certitude (le lumbago, par exemple). Et parfois, même la symptomatologie est imprécise. C'est le cas des maladies «idiopathiques» comme le colon irritable ou de nombreuses affections mentales pour lesquelles la recherche physiopathologique n'a pas encore permis de mettre en évidence le mécanisme, et dont les définitions mêmes sont régulièrement revues.

Ce rapide examen des entités et constructions nosologiques actuelles permet de comprendre que le raisonnement causal et le raisonnement physiopathologique n'ont de pertinence que dans les cas, respectivement, d'entités aux causes et mécanismes démontrés. Ces entités n'étant pas, aujourd'hui encore, très nombreuses, il en résulte des critiques parfaitement justifiées de l'application des raisonnements causal et physiopathologique dans de nombreuses situations, dans lesquelles le raisonnement probabiliste – ou l'usage de règles de prédiction clinique, s'il en existe d'effectivement validées – devrait évidemment être préféré.

Tableau 1. Type d'entité nosologique selon les manifestations, lésions, mécanismes et causes des affections

Groupe 6	Groupe 5	Groupe 4	Groupe 3	Groupe 2	Groupe 1
Affection définie par :	Affection définie par :	Affection définie par :	Affection définie par :	Affection définie par :	Affection due à une
Simple trouble =====	Plusieurs signes ou syndrome =====	Plusieurs signes et/ou syndrome + évolution =====	Signes et/ou syndrome + évolution + Lésions et/ou Physio-pathologie =====	Signes et/ou syndrome + évolution + Lésions et/ou Physio-pathologie + Cause =====	Anomalie innée de l'ADN + permanence + Conséquences Cliniques Lésionnelles Physio-pathologiques

2 ⇨ Place centrale des inférences causales dans la démarche diagnostique tout au long de l'histoire de la médecine savante

Ce qui vient d'être dit sur l'objectif du diagnostic, reconnaître *la* maladie affectant *le* malade, et sa mise en œuvre contemporaine, pourrait laisser croire que le diagnostic s'imposait comme une évidence à la médecine et aux médecins. En fait, il n'en fut rien.

L'étude de la terminologie et de ses usages s'avère ici, comme bien souvent, très utile : le mot « diagnostic », qui vient du grec *diagnosis* (de « discerner » devenu en latin *dignosco*), n'a pas été utilisé largement dans le contexte médical avant le XVI^e siècle et ne s'est vraiment imposé qu'au XVIII^e siècle. Auparavant, c'était un terme et une notion proche – mais pas totalement équivalente – qui était utilisée en médecine : celle de « différencier », une notion notamment utilisée par Galien, qui différenciait – c'est-à-dire classait ou plutôt *divisait* les symptômes – les maladies et les causes sur le modèle bien connu alors des « arbres de Porphyre » (par exemple, les maladies étaient divisées en simples et composées ; les simples en inflammations et intempéries ; les intempéries en froides et chaudes, etc.). Il est essentiel de remarquer que Galien mettait à peu près sur le même plan pour l'analyse du cas les maladies (qui étaient pour lui soit le résultat d'une « diathèse » – ou affection de tout le corps –, soit d'une « lésion » d'une fonction ou d'un organe), les symptômes (manifestations apparaissant anormales au malade ou au médecin, pas forcément rattachables à une maladie – Galien en distinguait deux sortes : les communs et les propres) et enfin les « causes », à la fois des symptômes et des maladies, qui étaient considérées les unes après les autres, et dans un ordre croissant de précision. Cette conceptualisation peut paraître confuse, d'autant que les écrits galéniques sont parfois contradictoires entre eux, mais elle s'éclaire un peu quand on se souvient que Galien recourait très largement à la conception hippocratique *dynamique* de la maladie, ne considérant pas la maladie en tant que telle, mais *la maladie chez le malade*, et attachant une grande importance à l'analyse de la rupture d'équilibre ayant déterminé les troubles, avec la considération de l'état pathologique (présent) des sujets, de leur état d'équilibre (antérieur), et des causes ayant concouru à cette rupture. Pour le médecin, il fallait *connaître* plutôt que *reconnaître* (diagnostiquer) la maladie, grâce à une évaluation multidimensionnelle visant à établir « ce qui n'allait pas », mais aussi à déterminer comment l'équilibre pouvait être rétabli – une évaluation dans laquelle l'analyse des

causes était évidemment centrale pour l'acte thérapeutique qui devait ramener le sujet à l'équilibre, habituellement par une action sur les contraires des causes ayant conduit à l'état pathologique. Cette analyse multidimensionnelle des malades, pour étrange qu'elle paraisse en ce début de XXI^e siècle, ne manque pas complètement de pertinence, puisque aujourd'hui encore, il faut le rappeler, beaucoup de symptômes ne sont pas reliés à une maladie, beaucoup le sont à plusieurs, des lésions peuvent relever de plusieurs causes et des causes peuvent être isolées ou associées à plusieurs maladies – dont certaines d'ailleurs présentent un continuum avec le sujet en bonne santé... La conception *dynamique* de la maladie, dite aussi «idiosyncratique» puisqu'il n'y a pas de maladie en tant que telle, mais *une maladie chez un malade*, fut très prégnante dans la médecine occidentale au moins jusqu'au XVI^e siècle, par exemple chez des auteurs comme Jacques Dubois ou encore Jean Fernel, qui demandaient au médecin confronté à une situation pathologique donnée de différencier, classer ou diviser les symptômes, les maladies et les *causes*, avant de mettre en œuvre une thérapeutique⁸.

8. Jacques Dubois tenta, dans les années 1530-1540, une synthèse de la conception galénique dans son *Methodus sex librorum Galeni in differentiis et causis morborum et symptomatum in tabellas sex ordine*, suivi du *De signis omnibus medicis hoc est, salubribus, insalubribus, et neutris, commentarius omnino necessarius medico futuro* : il construisit, à destination des étudiants, des «tables de divisions» (sous forme d'arbres de Porphyre) présentant toutes les «différences» (sortes) de maladies et de causes, ainsi que des différences et causes des symptômes, à partir des traités *De differentiis et causis morborum et symptomatum*, puis des tables de tous les signes de santé, d'état neutre et de maladie, à partir cette fois des traités de l'*Ars medica* (principalement), des *Tempéraments*, des *Lieux affectés*, du *Pouls* et du *Commentaire au Pronostic* hippocratique. Il est à noter que Dubois ne donna qu'une place très restreinte au traité des *Lieux affectés* et ne mentionna pas la méthode localisatrice des troubles énoncée au chapitre 5 du Livre 1, contrairement à Argenterio (*De morbis*, 1558) et surtout à Fernel (*Pathologia*, 1554) qui donnèrent à cette méthode un rôle essentiel pour la *connaissance* de la maladie touchant le malade. Le «diagnostic fernelien» comportait *deux temps* : le premier consistait à «rechercher l'endroit où était le mal», en suivant une méthode très proche de celle énoncée au chapitre 5 du Livre 1 des *Lieux affectés*, et le second à «reconnaître la maladie et la cause contenante d'icelle». Les signes qui servaient à la recherche du lieu affecté étaient tirés des «excréments», de la recherche de la «fonction lésée», de la propriété

L'accent mis sur la nécessité première de *discerner entre les maladies* responsables des troubles pathologiques, de « diagnostiquer » au sens actuel du terme, ne s'est alourdi qu'avec le développement des conceptions *ontologiques* de la maladie, où les maladies furent conçues comme des entités *indépendantes* du malade, à partir du XVI^e et surtout du XVII^e siècle. Ce fut d'ailleurs davantage un *retour* des conceptions ontologiques dans la pensée médicale qu'un développement *de novo*.⁹ Il y eut des prémisses de ce retour dès le milieu du XVI^e siècle, chez Fernel, mais surtout chez Fracastor et chez Van Lom (ou Lommius¹⁰, un auteur hollandais qui fut certainement influencé en la matière par Celse, un auteur latin du I^{er} siècle, qui était assez original sur cette question). Les conceptions ontologiques se retrouvent aussi dans les théories des chimistes, mais elles ne se furent pleinement exprimées dans le cadre de la médecine savante qu'avec Thomas Sydenham à la fin du XVII^e siècle. Celui-ci, en effet, mentionna très clairement en 1676, dans son introduction des *Observationes Medicae*, les « espèces de maladie » et la nécessité de les distinguer exactement les unes des autres, à l'exemple des botanistes, afin de *donner aux malades les traitements les plus adaptés* :

Il faut réduire toutes les maladies à des espèces précises et déterminées, avec le même soin et la même exactitude que les botanistes ont fait dans leur traités sur les plantes. Car il se trouve des maladies qui, étant du même genre et de même nom, et outre cela, semblables en quelques

et de la situation de la douleur et des « accidents propres », et servaient aussi au second temps pour caractériser la maladie et la cause.

9. Sur ce point, lire les analyses de Mirko GRMEK (1997, p. 157 et sq.). (Sur la critique de la conception ontologique de la maladie, cf. LAMBERT 2009. *NdÉ.*)

10. Van Lom considérait que l'identification du genre de la maladie était le préalable nécessaire à sa prise en charge correcte. Les *Medicinalium observationum libri tres, quibus notae morborum omnium et quae de his possint haberi praesagia judiciumque roponuntur* (1560) comprenait une première partie consacrée à la « reconnaissance » (Van Lom utilisa le verbe « *animadvertere* ») des maladies générales (essentiellement les fièvres), une deuxième consacrée à la reconnaissance des maladies spécifiques des parties du corps (classées de la tête au pied, puis par processus pathologique : inflammation, gangrène, érysipèle, etc.), et une troisième aux signes pronostiques. Sur cette œuvre, cf. DUFFIN (2006).

symptômes, sont néanmoins d'une nature bien différente, et demandent aussi un traitement différent. (Sydenham 1676.)

L'œuvre de Sydenham marqua certainement un tournant dans la conceptualisation des maladies et du diagnostic, et elle marqua un tournant dans la recherche nosologique qui connut un essor considérable au XVIII^e siècle, et dont l'œuvre bien connue de Boissier de Sauvages, publiée en 1732 et 1771, constitue assurément le sommet incontesté. Son titre, *Nosologia methodica, sistens morborum classes, genera et species juxta Sydenhami mentem et botanicorum ordinem*, est particulièrement éclairante de la dette de Boissier envers Sydenham. Mais si avec Boissier de Sauvages, la nosologie était voulue d'abord descriptive et «botanique», et donnait un rôle secondaire aux causes des maladies, ces dernières réapparurent au premier plan chez Pinel dans sa *Nosographie philosophique* publiée en 1798: «Il y avait besoin d'ordonner les maladies selon des "lois"», insista Pinel, qui les distribua en effet dans les classes des fièvres, des phlegmasies, des hémorragies, des névroses, des maladies dont le siège était dans le système lymphatique et celles dont le siège n'était pas déterminé...¹¹

L'évocation de Pinel permet de faire la transition avec le diagnostic clinique qu'on pourrait qualifier de «classique», qui se mit progressivement en place au cours du XIX^e siècle et qui resta enseigné sinon pratiqué jusqu'à aujourd'hui, sous la forme de la fameuse triade «diagnostic positif, diagnostic étiologique, diagnostic différentiel»:

- le diagnostic positif ou syndromique, où il s'agit de reconnaître le syndrome – un ensemble de symptômes et signes constituant une entité reconnaissable soit par l'uniformité de l'association stéréotypée des manifestations morbides, soit par le fait qu'elle traduit l'atteinte d'un organe ou d'un système bien défini. Par exemple: le syndrome méningé, le syndrome de diarrhée fébrile, le syndrome de polyarthrite aiguë, le syndrome cérébelleux, le syndrome pyramidal, etc.;
- le diagnostic étiologique, où il s'agit de trouver la cause vraisemblable du syndrome chez le malade, ce qui fait souvent envisager les

11. Il est tout à fait significatif qu'après Pinel, toutes les nosologies puis les classifications des maladies – jusqu'aux dernières révisions de la classification internationale des maladies (CIM9 et CIM10) – furent mixtes ou mêmes bigarrées, avec des symptômes, des syndromes et des maladies à mécanismes et à causes connus ou non...

grandes variétés de causes: les traumatiques, toxiques, nutritionnelles, infectieuses, vasculaires, tumorales, dysimmunitaires, dégénératives, génétiques ou congénitales et idiopathiques¹², et enfin;

- le diagnostic différentiel, où il s'agit de procéder à l'élimination des affections ressemblantes.

Si l'on reprend l'exemple donné plus haut (un sujet présentant des troubles de la marche associés avec un déficit moteur d'un membre inférieur, pour lequel on trouve à l'examen un syndrome pyramidal du même côté et un déficit de la sensibilité douloureuse de l'autre côté, et quelques troubles sphinctériens), le tableau présenté évoque fortement le diagnostic syndromique d'atteinte médullaire, et le diagnostic étiologique consistera à rechercher une compression mécanique, tumorale, une cause vasculaire (un angiome, par exemple), une affection malformative (comme la syringomyélie), puis le diagnostic différentiel conduira à éliminer les affections ressemblantes comme la compression nerveuse périphérique (mais il n'y aurait pas de syndrome pyramidal), la polyradiculonévrite (mais elle est habituellement bilatérale), la sclérose en plaques (mais il y aurait d'autres signes et une histoire de déficits régressifs), la conversion hystérique (mais la systématisation des troubles n'est pas usuelle), etc.

On aura remarqué que le diagnostic clinique «classique», comme le diagnostic ou plutôt l'évaluation galénique, accorde une place centrale à la recherche des causes: pendant presque toute l'histoire de la médecine pratique savante, cette recherche a été valorisée – on pourrait dire *paradoxalement*, puisque que les causes des maladies n'étaient en général pas précisément connues... Mais il faut cependant préciser qu'il s'agissait d'une causalité large, complexe, à plusieurs niveaux et multiple¹³: si les causes éloignées des maladies étaient souvent considérées

12. Les variétés «congénitales», «dégénératives» et «idiopathiques» ne peuvent bien entendu être mises au même niveau que les autres «causes» dans cette catégorisation avant tout opérationnelle pour la pratique clinique.

13. Les quatre causes d'Aristote (les «causes matérielles»: de quoi les choses étaient faites; les «causes formelles»: les «idées des choses» ou les modèles suivant lesquels elles avaient été faites; les «causes motrices»: celles qui avaient fait les choses; et les «causes finales»: les états finaux ou achevés des choses) furent largement utilisées par la médecine médiévale et encore au début de l'époque moderne, mais c'était surtout les trois catégories de

comme inaccessibles, les mécanismes – qu'on qualifie aujourd'hui de «physiopathologiques» –, eux, étaient accessibles et utilisables pour le raisonnement diagnostique, ou du moins le pensait-on.

3 ⇨ Les préoccupations causales fondatrices de la médecine pratique savante

La place centrale de la recherche des causes dans la «démarche diagnostique¹⁴» tout au long de l'histoire de la médecine savante est apparue, nous l'espérons, assez clairement. Nous voudrions terminer cette brève étude en rappelant que les préoccupations causales dans cette démarche furent même *fondatrices* de la médecine savante ou rationnelle. Les grands textes «constitutifs» du corpus hippocratique, puis de Galien, insistent en effet très fortement sur l'exigence de cette recherche des causes – causes naturelles bien sûr – à l'origine des troubles présentés par les malades, *afin de mieux les prendre en charge*. Deux textes hippocratiques, *Vents* et *l'Ancienne médecine*, le second probablement de la main d'Hippocrate même, sont particulièrement importants à considérer :

Un des points où l'on s'égaré est la question de savoir quelle peut être la cause des maladies, et quelles sont l'origine et la source des maux qui affligent le corps. En effet, si l'on connaissait la cause de la maladie, on serait en état d'administrer ce qui est utile, prenant dans les contraires l'indication des remèdes. De fait cette médecine est toute naturelle. (HIPPOCRATE, *Vents* 1, traduction Littré.)

Ainsi je crois fermement que tout médecin doit étudier la nature humaine, et rechercher soigneusement, s'il veut remplir ses obligations, quels sont les rapports de l'homme avec ses aliments, avec ses boissons, avec tout son genre de vie, et quelles influences chaque chose exerce sur chacun. Et il ne suffit pas de savoir simplement que le fromage est un mauvais aliment, parce qu'il cause des douleurs à ceux qui s'en rasasient; mais il faut savoir quelle douleur il cause, pour quelle raison,

causes du système galénique que les médecins utilisaient : les causes «antécédentes», qui correspondaient schématiquement aux facteurs internes prédisposant à la survenue de la maladie ; les causes «initiales», habituellement externes, qui déclenchaient la maladie ; et les causes «cohésives», qui s'apparentaient au processus physiopathologique lui-même.

14. Il s'agit bien sûr de la démarche diagnostique au sens large, définie dans la précédente section.

et à quelle humeur du corps il est contraire. Il y a en effet beaucoup d'autres aliments et boissons qui sont nuisibles à l'économie humaine, mais qui ne l'affectent pas de la même manière. Soit pour exemple le vin pur, qui, bu en grande quantité, jette l'homme dans une certaine faiblesse; on aura qu'à ouvrir les yeux pour connaître que la cause de cette faiblesse est dans la propriété du vin et dans le vin lui-même; et nous savons sur quoi, dans l'économie humaine, il porte son action. Cette vérité, qui est manifeste ici, je veux qu'elle le soit aussi dans les autres cas. (HIPPOCRATE, *Ancienne médecine* 20, traduction Littré.)

Cette exigence de la recherche des causes n'allait pas toutefois sans créer des tensions très fortes, voire des sensations de malaise chez les promoteurs mêmes de la médecine rationnelle, qui étaient bien conscients de la difficulté, voire de l'impossibilité à mener à bien cette recherche dans de nombreux cas. Un autre texte hippocratique (un passage de l'*Art*) illustre parfaitement le malaise des médecins rationnels qui étaient en pratique régulièrement confrontés à l'échec de la recherche des causes et finalement à l'échec de la prise en charge du malade – notamment dans le cas des *maladies internes* :

Dans ces cas, pour connaître, il faut bien plus de peine et bien plus de temps que si l'on employait les yeux. Ce qui échappe à la vue du corps est saisi par la vue de l'esprit; et les accidents qu'éprouve le malade dans ce retard sont imputables, non à celui qui le traite, mais à la constitution du patient et à la nature du mal. En effet le médecin, n'ayant pu connaître ni par la vue directe ni par les détails communiqués, la recherche par le raisonnement. Et de fait, les renseignements que les individus atteints de maladies cachées essayent de donner au médecin, sont dictés plus par les opinions que par une connaissance positive; car, s'ils avaient eu cette connaissance, ils ne seraient pas tombés malades, vu que c'est un savoir de même ordre de pénétrer la cause des maladies et d'être habile à y appliquer tous les traitements qui les empêchent de grandir. Donc, lorsque les renseignements ne peuvent fournir rien de précis et de certain, le médecin doit tourner ailleurs ses regards; et une telle lenteur est imputable, non à l'art, mais à la nature des corps malades. L'art attend, pour se mettre à l'œuvre, qu'il se soit rendu compte du mal, visant à le traiter plutôt avec prudence qu'avec témérité, avec douceur plutôt qu'avec violence. La nature, si elle donne le temps de pénétrer le mal, donnera aussi le temps de le guérir; mais si elle est vaincue dans l'intervalle que dure l'examen, soit parce que le secours du médecin a été tardivement réclamé, soit à cause de la rapidité du mal, l'issue sera funeste. (HIPPOCRATE, *Art* 11, traduction Littré.)

Entretenu par les échecs répétés – auxquels la médecine n'a cessé d'être confrontée –, ce malaise, qui était aussi perceptible dans d'autres

textes du corpus hippocratique du v^e siècle, s'est exacerbé au cours des siècles suivants et contribua vraisemblablement à donner naissance à plusieurs courants de pensée médicale, à des «sectes» comme on disait alors de manière non péjorative, qui se définirent et s'opposèrent précisément sur la question des causes des maladies et sur la nécessité faite au médecin de les connaître et de les rechercher. Ces sectes nous sont connues grâce à la préface du *De Medicina* de Celse et au traité *Des sectes aux étudiants* de Galien – un texte majeur qui a été lu, relu, médité, commenté, enseigné aux jeunes médecins jusqu'au xviii^e siècle. Dans celui-ci, Galien discutait les méthodes des trois sectes pour trouver des remèdes, et s'en prenait aux méthodiques et aux empiriques, ces derniers ne jurant – d'après lui – que par les causes évidentes et ce qu'il appelait le *théorème* du souvenir – répéter ce qui a déjà marché plusieurs fois.

Mais il existe une troisième espèce d'expérience qui est l'imitative et qui a lieu quand on expérimente à diverses reprises, dans des affections identiques, des moyens quelconques qui ont nui ou soulagé soit naturellement, soit par le hasard, soit qu'on ait essayé d'y avoir recours de propos délibéré. C'est là, suivant les empiriques, ce qui constitue surtout l'art; car c'est après avoir imité, non seulement deux ou trois fois, mais très souvent, le traitement qui a soulagé une première fois, c'est après avoir constaté ensuite que le plus ordinairement il produit les mêmes effets dans les mêmes affections, qu'ils appliquent le nom de *théorème* à ce souvenir (c'est à dire à l'ensemble) de tous ces cas, et qu'ils pensent être arrivés à un résultat digne de confiance et posséder une partie de l'art. Lorsqu'on eut ainsi rassemblé un grand nombre de ces théorèmes, la médecine fut constituée, selon eux, par leur réunion totale, et celui qui les réunit fut médecin. (GALIEN, *Des sectes aux étudiants*, chapitre 2, «Méthode des empiriques pour trouver les remèdes», traduction Daremberg.)

Galien prônait quant à lui la recherche des causes, afin que, celle-ci étant connues, il fût possible de traiter par les contraires. Dans le même texte, un peu plus loin, il apostrophait directement les empiriques sur l'exemple de deux malades mordus par un chien enragé:

Si, comme je l'ai également entendu au début de cette discussion, vous répétez: Tout ce qui est obscur est inutile, et si vous vous en tenez aux choses évidentes, comme vous l'avez dit au début, je vous montrerai peut-être ce que vous avez négligé, en vous remémorant les phénomènes: Deux hommes mordus par un chien enragé vont trouver leur médecin habituel pour réclamer des soins; chez tous les deux la bles-

sure est petite, de sorte que la peau même n'est pas tout à fait divisée : l'un des médecins traite seulement la blessure, ne s'inquiétant pas d'autre chose, et guérit la partie en peu de jours ; l'autre au contraire, lorsqu'il apprend que le chien était enragé, loin de s'empresse de faire cicatriser la plaie, l'agrandit de plus en plus, en employant pendant longtemps des médicaments forts et âcres ; et il oblige le malade à boire pendant longtemps des contre-poisons et des remèdes propres à guérir la rage, comme il les appelait lui-même. Or, voilà ce qu'il advint finalement au deux malades : l'un fut sauvé et guéri, c'est-à-dire celui qui avait bu le contre-poison ; l'autre persuadé qu'il n'avait aucun mal, fut subitement pris d'hydrophobie et mourut dans les convulsions. Croyez-vous que, dans ce cas, on ait cherché en vain à cause procatartique, et l'homme mourut-il par une autre raison que par la négligence du médecin, qui ne s'enquit nullement de la cause et n'employa pas le traitement tiré de l'observation ; il me semble que c'est là véritablement ce qui entraîna la mort du malade. (*Op. cit.*, chapitre 8, « Réfutation des méthodiques par les empiriques », traduction Daremberg.)

Ce (trop) bref retour aux premiers siècles d'existence de la médecine savante a permis de constater l'ancienneté et la permanence des tensions existant autour des approches « déterministes » utilisées en pratique médicale, notamment dans la conduite de la démarche diagnostique. Ces tensions suscitérent précocement des approches alternatives, qualifiées d'« empiriques » – fondées sur la « mémoire seule », disait Galien – mais que l'on pourrait aisément aujourd'hui qualifier de pré- ou cryptoprobabilistes et qui se maintinrent (sans être toujours marginales) en médecine tout au long de son histoire, jusqu'à aujourd'hui où elles s'expriment sous la forme, à peine plus structurée, des règles de prédiction clinique¹⁵. On pourrait d'ailleurs considérer que ces tensions sont *inhérentes à la médecine*, à la fois pratique et science, exigence de guérir ou du moins de soulager et quête de la compréhension du vivant et de ses pathologies, et dont les praticiens sont quotidiennement confrontés à la fois aux échecs d'efficacité et aux difficultés de répondre de manière satisfaisante aux questions des malades et des sociétés sur les causes des maladies. Il n'est pas douteux qu'elle continuera encore

15. Comment ne pas être frappé par la similitude de la collection « d'un grand nombre de théorèmes du souvenir » prônée par les empiriques antiques et le récent ouvrage de Mark EBELL (2001), *Evidence-Based Diagnosis. A Handbook of Clinical Prediction Rules*, qui présente lui aussi une collection de règles probabilistes développées pour diverses situations cliniques ?

longtemps, dans le domaine du diagnostic comme dans d'autres domaines de la pratique, à vouloir concilier le désir d'être avant tout efficace et celui d'expliquer, et de balancer entre les deux quand ces deux désirs ne peuvent être satisfaits conjointement¹⁶.

☉Références bibliographiques☉

B

- BALL G.V. (1971), "Two epidemics of gout", *Bulletin of the History of Medicine*, 45: 401-408.
BURNSIDE E. (2003), "Bayesian networks. Computer-assisted diagnosis support in radiology", *Academic Radiology*, 12: 422-430.

C

- COSTE J. & PAOLAGGI J.-B. (2006), *Le Raisonnement médical. De la science à la pratique clinique*, Estem.

D

- DUFFIN J., "Jodocus Lomnius's little golden book and the history of diagnostic semiology", *Journal of the History of Medicine and Allied Sciences*, 61: 249-287.

E

- EBELL M. (2001), *Evidence-Based Diagnosis. A Handbook of Clinical Prediction Rules*, Springer.

G

- GARROD A.B. (1859), *The Nature and Treatment of Gout and Rheumatic Gout*, Walton and Moberly.

- GRMEK M. (1997), *Histoire de la pensée médicale*, vol. 2, Seuil.

L

- LAMBERT G. (2009), *Vérole, cancer & Cie. La société des maladies*, Seuil.

R

- RANSOHOFF D.F. & FEINSTEIN A.R. (1976), "Editorial: Is decision analysis useful in clinical medicine?", *Yale Journal of Biology and Medicine*, 49: 165-168.

S

- SCHWARTZ W.B., WOLFE H.J. & PAUKER S.G. (1981), "Pathology and probabilities: a new approach to interpreting and reporting biopsies", *New England Journal of Medicine*, 305: 917-923.

- STIELL I.G., MCKNIGHT R.D., GREENBERG G.H., McDOWELL I., NAIR R.C., WELLS G.A., JOHNS C. & WORTHINGTON J.R. (1994), "Implementation of the Ottawa ankle rules", *JAMA*, 271: 827-832.

- SYDENHAM T. (1676), *Observationes medicae*, London. Trad. de Jault sous le titre *Médecine pratique de Sydenham*, Paris, 1774.

16. Ces réflexions conclusives ne sont pas sans rappeler certains propos de Steeves Demazeux (ce volume), dans le contexte assez spécifique de la médecine psychiatrique. (NdÉ.)

T

TVERSKY A. & KAHNEMAN D. (1974), “Judgment under Uncertainty: Heuristics and Biases”, *Science*, 185: 1124-1131.

↻ Notice biographique ↻

Joël Coste est professeur des universités et praticien hospitalier à l'Université Paris Descartes et à l'hôpital Cochin, directeur d'études à l'École pratique des hautes études (IV^e Section). Ses domaines de recherche incluent la métrologie médicale, la mesure de la qualité de vie liée à la santé, l'évaluation des pratiques médicales, la biopathologie historique, l'épidémiologie historique et l'histoire culturelle des maladies. Il est l'auteur ou le coauteur de 225 publications médicales et de 7 livres, dont *Le Raisonnement médical. De la science à la pratique clinique* (2001, avec J.-B. Paolaggi).

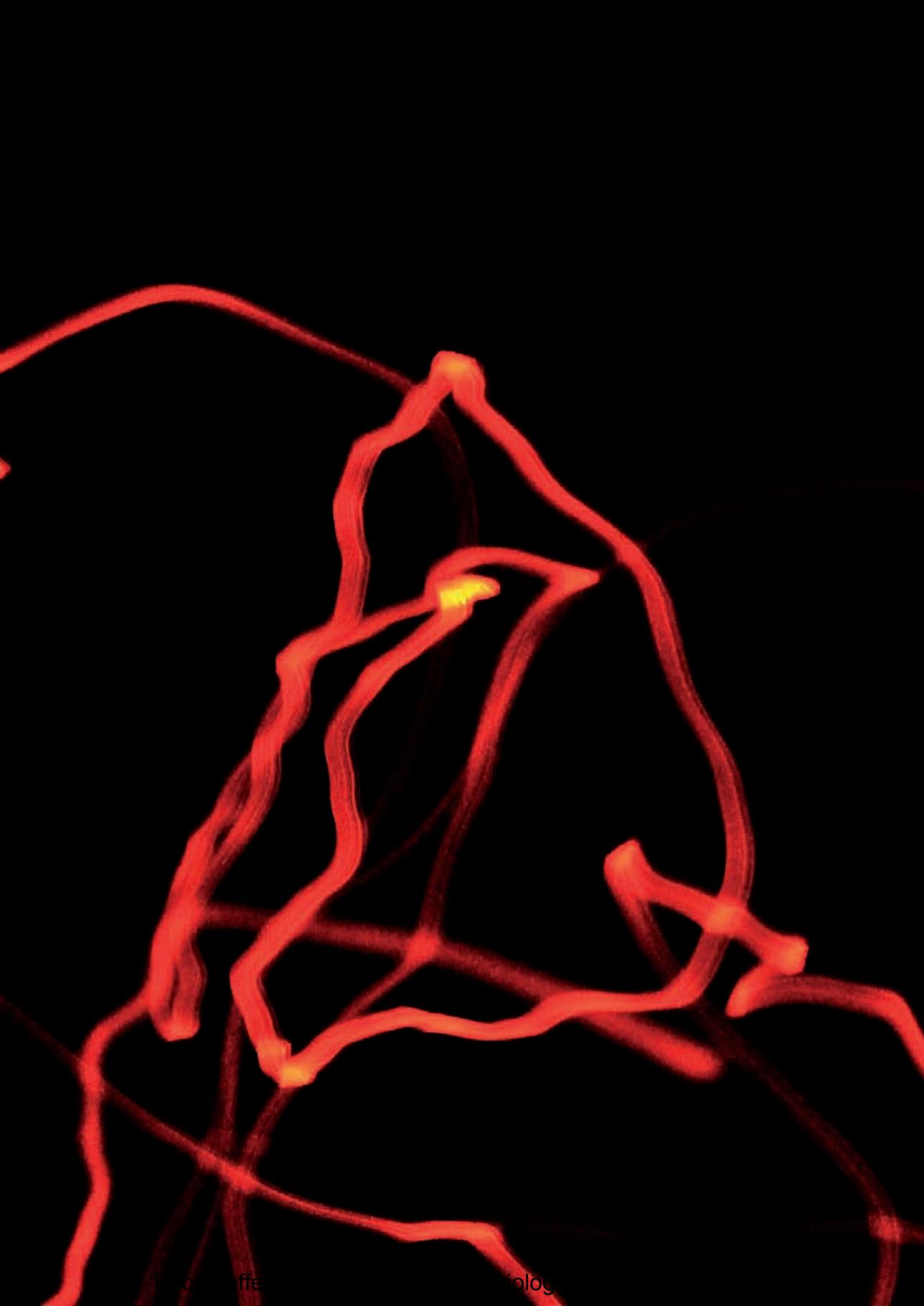
Avertissement. Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur et des auteurs. Les textes et illustrations figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage personnel, scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document selon ce modèle :

Joël Coste, «Inférences causales et probabilistes dans le raisonnement diagnostique: un éclairage historique sur le débat contemporain», in Gérard Lambert et Marc Silberstein (dir.), *Matière première. Revue d'épistémologie [en ligne]. Nouvelle série, N° 1/2010 : Épistémologie de la médecine et de la santé*, Éditions Matériologiques. Mis en ligne le 7 octobre 2010.
URL: www.materiologiques.com

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.



Varia



Michael Esfeld **Les fondements de la causalité**

Bertrand Russell soutient dans un article célèbre de 1912 que la notion de causalité est dépassée (RUSSELL 1912 [2006]). Selon lui, il n'est plus question de causalité dans les théories physiques fondamentales du xx^e siècle. En bannissant la notion de causalité, Russell ne veut pas dire que les énoncés causaux communs sont faux – des énoncés comme, par exemple, «Le mal de tête de Marie a disparu parce qu'elle a pris un comprimé d'aspirine», «Jean est mort du cancer du poumon parce qu'il a fumé toute sa vie», «Un ouragan a causé la destruction de La Nouvelle-Orléans», «Les gènes exercent une influence causale sur le comportement des organismes», «Pierre prend le TGV parce qu'il pense que c'est le moyen de transport le plus rapide pour se rendre de Paris à Lyon», etc. Russell maintient que si nous tenions compte des faits physiques grâce auxquels de tels énoncés sont vrais, nous ne découvririons pas de connexions réelles de cause à effet dans le monde. Des énoncés causaux comme ceux que nous venons de mentionner sont vrais non parce qu'il existe des relations réelles physiques de cause à effet mais parce que le monde exhibe certains traits physiques non causaux que nous pouvons conceptualiser sous la forme d'énoncés causaux. En bref, la causalité n'existe pas dans le monde, mais il y a des vérifacteurs¹ non causaux pour des énoncés causaux.

1. Si une proposition portant sur le monde est vraie, il y a un élément dans le monde réel grâce auquel la proposition en question est vraie ; on appelle cet élément «vérifacteur».

Depuis Russell jusqu'à aujourd'hui, le débat philosophique sur les fondements de la causalité tourne autour de la question suivante : la causalité est-elle un trait fondamental du monde ? Ou s'agit-il de quelque chose qui est dérivé d'autres traits fondamentaux du monde ? Dans cet article, nous allons retracer l'état actuel du débat. Nous commencerons par les théories de la causalité dans la tradition de Russell, c'est-à-dire les théories de la causalité comme régularité (section 1), et nous jetterons un bref coup d'œil sur les théories de la causalité comme processus physique (section 2). Puis, nous considérerons les arguments récents en faveur des pouvoirs causaux, c'est-à-dire des arguments qui militent en faveur de la conception de la causalité comme trait fondamental du monde (section 3). Après ce tour d'horizon des positions philosophiques concernant la causalité, nous tiendrons compte de la causalité en physique (section 4) ainsi que dans les sciences spéciales comme la biologie et la psychologie (section 5), argumentant, contre Russell, que ces sciences notamment suggèrent un engagement ontologique en faveur des pouvoirs causaux. Cet engagement ontologique ouvre la perspective d'un matérialisme et d'un réductionnisme non-éliminativistes qui gardent les traits caractéristiques des propriétés biologiques, mentales, etc., et les intègrent en même temps dans le monde physique, matériel.

1 ↻ La causalité comme régularité

La critique de Russell de la notion de causalité s'insère dans une tradition qui commence avec David Hume. Selon Hume, nous concevons l'existence d'un lien nécessaire entre deux événements si nous considérons l'un comme la cause et l'autre comme son effet : l'événement qui est la cause engendre l'événement qui est l'effet et est ainsi la raison de son existence. Toutefois, d'après Hume, tout ce que nous observons est une succession temporelle d'événements contigus dans l'espace : par exemple, une boule de billard touche une autre boule de billard, celle-ci se met en mouvement, et la première arrête son mouvement.

Ce qu'on entend aujourd'hui par *métaphysique humienne*, c'est la position selon laquelle il n'y a pas de faits modaux – comme des connexions nécessaires – dans le monde. C'est également la position que défend Russell dans sa critique de la notion de causalité. Le philosophe qui a contribué le plus au développement de la métaphysique humienne dans la deuxième moitié du xx^e siècle est David Lewis. D'après Lewis, tout ce qui existe dans le monde, ce sont (a) des relations métriques

entre des points de l'espace-temps et (b) des propriétés physiques fondamentales qui se situent en ces points (cf. LEWIS 1986a, introduction; LEWIS 1994). Tout le reste de ce qui existe dans le monde est fixé par la distribution des propriétés physiques fondamentales en des points de l'espace-temps, dans le sens où il s'agit d'un trait de cette distribution. Par exemple, un organisme comme une plante ou un être humain n'est rien d'autre que la configuration de propriétés physiques fondamentales instanciée en une certaine région de l'espace-temps quadridimensionnel. Cette position est évidemment un matérialisme basé sur les théories physiques du xx^e siècle; la question est toutefois de savoir si ce réductionnisme austère est capable de tenir compte des traits caractéristiques des organismes et surtout ceux des êtres humains. Nous reviendrons sur cette question dans la section 5.

La métaphysique humienne n'a pas besoin de la causalité comme ciment qui unit le monde. Le filet constitué par les relations métriques entre des points de l'espace-temps est suffisant pour unir le monde. Les propriétés physiques fondamentales en des points de l'espace-temps sont des propriétés *intrinsèques*. Chacune de ces propriétés appartient à un point de l'espace-temps – ou à une particule localisée en un point – indépendamment de l'existence de quoi que ce soit d'autre dans le monde. Par exemple, la propriété de Pierre d'avoir une masse de 80 kilos peut postuler au titre de propriété intrinsèque macroscopique. Par contre, la propriété de Pierre d'être plus lourd que Paul n'est pas une propriété intrinsèque, mais consiste en une relation entre Pierre et Paul. Une relation du type «*x* est plus lourd que *y*» n'est pas fondamentale: elle est fixée par les masses que *x* et *y* (Pierre et Paul) possèdent l'un indépendamment de l'autre. D'après Lewis, il en va de même pour toutes les relations, sauf les relations métriques entre les points de l'espace-temps: les relations qui existent dans le monde sont fixées par la distribution des propriétés physiques fondamentales et intrinsèques en des points de l'espace-temps.

De plus, les propriétés physiques fondamentales sont des propriétés *catégoriques*: elles sont des qualités pures. Autrement dit, en tant que telles, elles ne possèdent pas la disposition de causer quoi que ce soit. C'est la raison pour laquelle il n'y a pas de connexions nécessaires dans la nature: aucune des propriétés physiques ne possède le pouvoir (disposition causale) d'engendrer d'autres propriétés. Par conséquent, chaque occurrence d'une propriété physique fondamentale est contin-

gente : on peut tenir comme inchangée n'importe quelle occurrence d'une propriété physique fondamentale en un point de l'espace-temps quand bien même toutes les autres occurrences de propriétés auraient changé. Il faut dès lors reconnaître la distribution entière des propriétés physiques fondamentales comme primitive, c'est-à-dire comme quelque chose qu'il faut accepter comme point de départ ne possédant pas d'explication. Néanmoins, la métaphysique humienne est économe : étant donné les relations métriques entre les points de l'espace-temps et les propriétés physiques fondamentales, intrinsèques et catégoriques qui se situent en ces points, tout ce qui existe dans le monde est fixé et peut être expliqué.

La distribution des propriétés physiques fondamentales dans l'espace-temps manifeste des régularités : certaines propriétés se produisent toujours – ou presque toujours – conjointement. Certaines de ces régularités sont, selon Lewis, des lois de la nature, à savoir celles qui figurent comme axiomes dans la description du monde qui réalise le meilleur équilibre entre simplicité et contenu empirique. Les lois de la nature sont ainsi elles aussi fixées par la distribution des propriétés physiques fondamentales dans tout l'espace-temps, et elles sont contingentes (cf. LEWIS 1973a, p. 72-75).

Il en va de même pour les relations de causalité. Les causes et les effets sont des événements, et il convient d'adopter une conception précise de ce que l'on entend par «événement» : un événement est l'occurrence d'une propriété, à savoir, en fin de compte, l'occurrence d'une propriété physique fondamentale en un point de l'espace-temps. D'après la *théorie de la causalité comme régularité*, certaines régularités dans la distribution des propriétés physiques dans le monde sont des relations causales. Hume pose trois conditions nécessaires et qui, ensemble, sont suffisantes pour que la relation entre un événement e_1 et un autre événement e_2 soit une relation causale :

- 1) e_1 a lieu avant e_2 .
- 2) e_1 est spatio-temporellement contigu avec e_2 .
- 3) Toujours quand se produit un événement du même type qu' e_1 , il y a un autre événement du même type qu' e_2 qui succède au premier événement et qui est spatio-temporellement contigu avec lui (cf. HUME 1739, livre I, part. III, HUME 1748, section VII).

Par conséquent, la réponse à la question de savoir si oui ou non deux événements sont liés par une relation de cause à effet dépend de ce qui

a lieu ailleurs dans le monde. Considérons un événement e_1 qui est une occurrence de la propriété F et un événement e_2 qui est une occurrence de la propriété G . Supposons que e_1 et e_2 se succèdent et soient spatio-temporellement contigus. D'après l'analyse de la causalité en termes de régularités, la réponse à la question de savoir si e_1 cause e_2 dépend du fait que les événements qui sont F et les événements qui sont G se succèdent oui ou non de façon régulière. En d'autres termes, e_1 cause e_2 si et seulement si, dans des circonstances similaires, tous les F sont suivis par des G – en bref, si et seulement si c'est une loi de la nature que les F et les G se succèdent. La causalité se réduit ainsi à la succession régulière d'événements de mêmes types. À l'époque contemporaine, John Mackie notamment a essayé de formuler un critère général qui précise la conception de Hume, combinant des conditions nécessaires et suffisantes (cf. MACKIE 1965 et 1974, chap. 3). Ses efforts sont aujourd'hui poursuivis par Gerd Grasshoff et ses collaborateurs (cf. notamment GRASSHOFF & MAY 2001).

David Lewis, en revanche, ne soutient pas une théorie de la causalité comme régularité à la Hume mais une théorie contrefactuelle de la causalité. En bref, il y a un rapport de dépendance causale entre deux événements e_1 et e_2 si et seulement si la condition suivante est satisfaite : si e_1 avait (n'avait pas) eu lieu, e_2 aurait eu lieu également (n'aurait pas eu lieu non plus) (cf. LEWIS 1973b et 2004). Par exemple, si Pierre n'avait pas fumé toute sa vie, il ne serait pas mort du cancer du poumon ; si les réacteurs étaient tombés en panne, l'avion se serait écrasé. La question est cependant de savoir quels sont les vérificateurs de tels énoncés. Lewis ne peut pas soutenir que ces énoncés sont vrais en vertu d'une connexion réelle entre les événements en question qui se base sur des pouvoirs causaux. Selon lui, c'est la distribution des propriétés physiques fondamentales dans l'espace-temps en entier – et notamment les régularités dans cette distribution – qui rend vrai des énoncés contrefactuels portant sur le monde réel. Lewis introduit la notion d'autres « mondes possibles » pour établir la valeur de vérité des énoncés contrefactuels : ceux-ci sont arrangés en fonction de leur distance au monde actuel et permettent ainsi d'évaluer les énoncés contrefactuels. Lewis adopte une attitude réaliste envers ces mondes possibles (cf. LEWIS 1986b). Sa théorie contrefactuelle de la causalité ne condamne pourtant pas à accepter cet engagement ontologique concernant de tels mondes possibles : ce qui rend vrai les énoncés contrefactuels par rapport au monde

réel, c'est la distribution des propriétés physiques fondamentales dans le monde réel (cf. LOEWER 2007, p. 308-316).

La théorie contrefactuelle de la causalité s'avère ainsi être une version sophistiquée de la théorie de la causalité comme régularité : le fait qu'il y ait ou non un rapport causal entre deux événements e_1 et e_2 ne dépend pas du fait que oui ou non il y a beaucoup d'événements du même type qu' e_1 auxquels succèdent des événements du même type qu' e_2 . Si oui ou non il y a un rapport causal entre deux événements dépend notamment des lois de la nature, et celles-ci sont à leur tour fixées par la distribution entière des propriétés physiques fondamentales dans le monde.

La théorie de la causalité comme régularité, y compris la théorie contrefactuelle, se base dès lors sur les deux présuppositions suivantes :

1) *La causalité comme relation extrinsèque* : si oui ou non il y a un rapport causal entre deux événements dépend de la distribution des propriétés physiques fondamentales dans le monde en entier.

2) *La causalité comme relation contingente* : l'événement qui est la cause ne produit pas l'existence de l'événement qui est l'effet. Il n'existe pas de pouvoirs causaux dans le monde : les propriétés ne sont pas causales en tant que telles mais catégoriques. Il faut considérer la distribution entière des propriétés physiques fondamentales comme point de départ ; sur la base de cette distribution comme fait primitif, des régularités – et à travers elles des relations causales – sont fixées.

La question centrale pour l'évaluation de la théorie de la causalité comme régularité est de savoir si ces deux présuppositions sont convaincantes. On les examinera dans la suite de l'article.

2 ↻ La causalité comme processus physique

Une position importante qui cherche à transformer la conception humienne de la causalité est la théorie de la causalité comme processus physique – plus précisément comme processus de transfert ou d'échange d'une grandeur physique conservée comme de l'énergie. Cette théorie a été développée notamment par Wesley Salmon, Phil Dowe et Max Kistler². Dowe introduit la distinction importante suivante : un proces-

2. Cf. SALMON (1998, chap. 1, 12, 16, 18), DOWE (2000, chap. 3 à 5), KISTLER (1999, chap. 1.4).

sus causal est une ligne de monde tracée par un objet qui possède une grandeur physique conservée ; une interaction causale est une intersection de lignes de monde qui implique l'échange d'une grandeur physique conservée, notamment de l'énergie (DOWE 2000, p. 90).

La théorie de la causalité comme transfert ou échange d'une grandeur physique conservée contredit la première présupposition de la conception humienne : la causalité n'est pas une relation extrinsèque mais une relation intrinsèque, dépendant uniquement des propriétés des *relata*, indépendamment de ce qui se passe ailleurs dans le monde. Plus précisément, le fait qu'il y ait ou pas un rapport causal entre deux événements spatio-temporellement contigus dépend uniquement de ce qui a lieu dans la région spatio-temporelle qu'occupent ces deux événements, à savoir si oui ou non il y a un transfert ou échange d'une grandeur physique conservée entre les deux événements en question.

Cette conception de la causalité comme rapport local s'oriente vers la physique classique. Sa présupposition de localité n'est pas satisfaite en physique quantique, l'interaction avec un système quantique pouvant causer la manifestation de corrélations non locales qui ne se limitent donc pas à la région spatio-temporelle dans laquelle l'interaction en question a lieu. Elle est également en contradiction avec la théorie de la relativité générale : celle-ci ne permet pas de considérer l'énergie gravitationnelle comme étant localisée en des points ou de petites régions de l'espace-temps (cf. CURIEL 2000, LAM 2005). On peut bel et bien contester la présupposition humienne selon laquelle n'importe quel rapport causal entre deux événements dépend de ce qui a lieu dans le monde en entier, mais il semble qu'on ne peut pas remplacer cette présupposition simplement par une théorie locale de la causalité.

Bien que la théorie de la causalité comme processus physique contredise la première présupposition de la conception humienne, on peut tout de même l'interpréter comme tombant sous la métaphysique humienne car elle ne prend pas position quant à la deuxième présupposition. Or, c'est cette deuxième présupposition – à savoir la conception des propriétés physiques comme catégoriques et ainsi non modales – qui est au centre de ce qu'on entend aujourd'hui par la métaphysique humienne. Rien n'empêche dans le cadre de la théorie de la causalité comme processus physique de considérer les suites d'événements qui sont des processus causaux comme des occurrences de propriétés catégoriques, à

savoir comme des occurrences de propriétés non causales, ne possédant pas de pouvoirs causaux.

Toutefois, on peut également interpréter la théorie de la causalité comme processus physique de façon à ce qu'elle contredise non seulement la première mais encore la deuxième présupposition de la métaphysique humienne : on peut regarder les propriétés physiques – y compris les grandeurs conservés – comme des propriétés causales, c'est-à-dire comme des pouvoirs qui produisent l'existence de leurs effets. Quoi qu'il en soit, le point central du débat philosophique autour des fondements de la causalité ne concerne pas la question de savoir si oui ou non la causalité consiste en des processus physiques locaux mais la question de savoir si les propriétés physiques sont catégoriques ou causales (dans ce dernier cas, elles seraient des pouvoirs).

3 ↻ Les pouvoirs causaux

La théorie humienne de la causalité a dominé la philosophie du ^{xx}e siècle grâce à son lien avec l'empirisme logique. Néanmoins, à partir des années 1980 notamment, des objections graves contre cette théorie ont été conçues qui ont mené à l'élaboration d'une autre théorie de la causalité, considérant les propriétés physiques comme étant des propriétés causales en tant que telles. Si, comme le soutient la métaphysique humienne, les propriétés physiques sont des propriétés catégoriques, elles sont des qualités pures. Par conséquent, ce qu'est une propriété ne fixe pas les rapports causaux – et pas non plus les rapports nomologiques (les lois de la nature) – dans lesquels les occurrences de la propriété en question peuvent figurer. Ceux-ci ne sont pas fixés par le caractère qualitatif de la propriété en question mais ils dépendent des autres propriétés qui existent dans le monde. Les rapports causaux et nomologiques dans lesquels les occurrences d'un type de propriété donné figurent varient d'un monde possible à l'autre, dépendant de la distribution entière des propriétés physiques fondamentales dans le monde en question. Par exemple, il y a, selon cette position, un autre monde possible dans lequel la propriété qui joue le rôle de masse dans le monde réel y joue le rôle de charge, et *vice versa*.

La conception humienne des propriétés comme catégoriques mène ainsi aux deux conséquences suivantes :

(a) Comme le caractère qualitatif d'une propriété est indépendant de tout rapport causal et nomologique, il s'agit de quelque chose de

primitif, étant indéfinissable. Deux types de propriétés, étant instanciées dans deux mondes possibles, peuvent figurer dans exactement les mêmes rapports causaux et nomologiques dans ces deux mondes et, néanmoins, il peut s'agir de deux types de propriétés différents. Il peut donc y avoir une différence entre deux mondes possibles qui consiste uniquement en une différence du caractère qualitatif entre des propriétés qui ne se manifeste nulle part. Cette conséquence est connue dans la littérature sous le nom de «quiddité», ce concept voulant dire que le caractère qualitatif d'une propriété est un fait primitif, indéfinissable (cf. par ex. BIRD 2007, chap. 4.2).

(b) Nous n'avons aucun accès cognitif au caractère qualitatif des propriétés. Toutes nos descriptions ne peuvent spécifier ce qui existe dans le monde que jusqu'à l'équivalence causale-nomologique, car c'est uniquement par le biais des relations causales que nous pouvons prendre connaissance de ce qui existe dans le monde. Si le caractère qualitatif des propriétés est séparé des rapports causaux-nomologiques dans lesquels elles figurent, nous ne pouvons alors pas connaître ce caractère qualitatif. Cette conséquence est connue dans la littérature sous le nom d'«humilité».

La conséquence métaphysique de quiddité et la conséquence épistémologique d'humilité sont des conséquences incommodes : on postule qu'il existe quelque chose dans le monde, à savoir le caractère qualitatif des propriétés, qu'on ne peut, par principe, pas connaître et qui, de plus, est indéfinissable, donnant lieu à des différences entre des mondes possibles qui ne font pas de différence car ne se manifestant nulle part. David Lewis lui-même a accepté les conséquences de quiddité et d'humilité dans l'un des derniers papiers qu'il a rédigé (LEWIS 2009), mais la mise en évidence de ces deux conséquences a plutôt provoqué la recherche d'une autre conception des propriétés.

L'idée est de dire que ce qu'est une propriété inclut la disposition de produire certains effets, de sorte que ses effets révèlent le caractère qualitatif de la propriété en question. On peut distinguer deux versions de cette idée dans la littérature contemporaine :

(1) La première version soutient que chaque propriété est à la fois causale-dispositionnelle et catégorique. La distinction entre «causal-dispositionnel» et «catégorique» n'est pas une opposition entre deux types

de propriétés: il s'agit de deux types de prédicats que nous utilisons pour décrire les mêmes propriétés³.

(2) L'autre version considère toutes les propriétés comme étant des pouvoirs. Chaque propriété est le pouvoir (disposition causale) de produire certains effets (cf. SHOEMAKER 1980, BIRD 2007).

Il n'y a pas de conflit substantiel entre ces deux versions. On peut dire qu'il s'agit plutôt d'accentuations différentes d'une seule et même position. La première version ne tient pas la distinction entre «causal-dispositionnel» et «catégorique» pour une opposition ontologique. On ne peut même pas parler de deux aspects différents des propriétés (car de tels aspects seraient à leur tour des propriétés). L'autre version ne conçoit pas les propriétés comme des potentialités pures. Les pouvoirs sont des propriétés réelles et actuelles. Ils sont certaines qualités, à savoir des pouvoirs de produire certains effets spécifiques. On peut dès lors résumer ces deux nuances de la même théorie des propriétés de la façon suivante: *dans la mesure où les propriétés sont certaines qualités, elles sont causales – c'est-à-dire qu'elles consistent en des pouvoirs de produire certains effets*. Prenons la charge comme exemple pour illustrer cette conception: dans la mesure où la charge est une propriété qualitative, distincte par exemple de la masse, elle est le pouvoir de créer un champ électromagnétique, pouvoir se manifestant par l'attraction des objets de charge opposée et la répulsion des objets de même charge.

Il est important de noter que les pouvoirs ne sont pas des propriétés additionnelles. L'adhérent à la métaphysique humienne et son adversaire sont d'accord au sujet des propriétés qui existent dans le monde. La distinction entre ces deux positions concerne l'ontologie des propriétés, c'est-à-dire la question de savoir si les propriétés sont catégoriques de sorte que le monde est une vaste mosaïque d'occurrences de propriétés sans liaisons entre elles ou si les propriétés sont causales de sorte que certaines propriétés produisent l'existence d'autres propriétés. La dispute porte sur les fondements de la causalité: d'après la théorie causale des propriétés, il y a un lien métaphysiquement nécessaire entre la présence de certaines propriétés et la production de certains effets car l'essence des propriétés n'est pas purement qualitative mais consiste en le pouvoir de produire certains effets. D'après la métaphy-

3. Cf. notamment MUMFORD (1998, chap. 9), HEIL (2003, chap. 11), KISTLER (2005).

sique humienne, par contre, il n'y a pas de lien réel de production entre cause et effet mais uniquement certaines régularités contingentes dans la distribution des propriétés qui rendent vrai des énoncés causaux.

La théorie causale des propriétés reconnaît dès lors comme primitif le fait qu'en tant qu'étant certaines qualités, les propriétés sont des pouvoirs de produire certains effets; mais cette théorie n'a pas besoin d'accepter la distribution entière des propriétés physiques fondamentales dans tout l'espace-temps comme un fait primitif qui ne peut pas être expliqué: si le monde est déterministe, on n'a besoin que d'accepter la distribution des propriétés physiques fondamentales à l'état initial du monde comme fait primitif; les pouvoirs inclus dans cette distribution engendrent tout le reste du monde. Si le monde n'est pas déterministe, il y a quand même des probabilités objectives fondées sur les pouvoirs causaux et ainsi des explications probabilistes du développement de la distribution des propriétés physiques fondamentales dans l'espace-temps.

Par conséquent, la vision causale des propriétés implique non seulement une théorie anti-humienne de la causalité mais également une théorie anti-humienne des lois de la nature: les lois de la nature sont métaphysiquement nécessaires au lieu d'être contingentes car elles dérivent de l'essence des propriétés. Il n'est pas possible que les rapports nomologiques dans lesquels figure un type de propriété donné changent d'un monde possible à l'autre, puisque c'est le caractère qualitatif de la propriété qui fixe les lois qui se rapportent à elle. Il n'y a dès lors pas de quiddités car ce que sont les propriétés consistent en la production de certains effets, et pas d'humilité car ce que sont les propriétés se manifeste dans les effets qu'elles produisent.

Si la raison principale pour adopter la théorie causale des propriétés est d'éviter les engagements ontologiques en faveur des quiddités et de l'humilité, on ne peut pas maintenir que les propriétés fondamentales ont besoin de conditions externes pour manifester les pouvoirs qu'elles sont: si tel était le cas, il pourrait y avoir des propriétés de deux types différents P et P^* présentes dans le monde sans que cette différence ne se manifeste nulle part, parce que les conditions permettant la manifestation des pouvoirs qui distinguent ces deux propriétés manquent toujours. Il ne pose pas de problème de concevoir les propriétés physiques fondamentales comme n'ayant pas besoin de conditions externes pour manifester les pouvoirs qu'elles sont, parce que ces pouvoirs ne

sont pas des potentialités pures mais des propriétés réelles, actuelles. Pour comprendre ce point, il ne faut pas se représenter des dispositions macroscopiques, comme celle de l'eau de dissoudre le sucre, mais plutôt des dispositions plus fondamentales comme celle des atomes radioactifs de se désintégrer spontanément ou celle des particules avec une charge de créer spontanément un champ électromagnétique.

Rien n'empêche que des propriétés causales puissent être des propriétés intrinsèques. Un objet peut posséder de telles propriétés indépendamment de l'existence de quoi que ce soit d'autre dans le monde. De plus, les effets qu'un objet ou qu'un événement produit en vertu de ses propriétés causales peuvent également être des propriétés intrinsèques : bien que ces propriétés existent dans le monde réel uniquement parce qu'elles sont produites par d'autres propriétés, de mêmes types de propriétés peuvent faire partie de l'état initial d'un autre monde.

4 ↻ La physique et la causalité

Comme nous l'avons mentionné au début de l'article, Russell soutient que les théories physiques du xx^e siècle excluent une conception de la causalité en termes de pouvoirs causaux. Ce jugement est partagé par nombre de philosophes de la physique aujourd'hui encore⁴. D'un autre côté, il y a des philosophes contemporains qui mettent en évidence que des physiciens utilisent un vocabulaire dispositionnel en décrivant les propriétés physiques fondamentales. Ils en tirent la conclusion que les théories physiques fondamentales contiennent un engagement ontologique envers des dispositions irréductibles, c'est-à-dire des pouvoirs causaux (cf. MUMFORD 2006, p. 475-477). Un simple regard sur les théories physiques fondamentales n'est donc pas suffisant pour trancher la dispute philosophique au sujet des fondements de la causalité.

Les théories physiques fondamentales contemporaines mettent l'accent sur des relations plutôt que sur des propriétés intrinsèques, tandis que le débat purement philosophique autour des fondements de la causalité tient comme admis que les propriétés physiques fondamentales sont des propriétés intrinsèques et se pose la question de savoir si celles-ci sont catégoriques ou causales. La théorie quantique décrit les objets physiques fondamentaux (comme par exemple les électrons,

4. Cf. notamment les contributions dans PRICE & CORRY (2007).

les protons, etc.) comme étant dans des états intriqués : au lieu d'avoir chacun des propriétés intrinsèques dont les valeurs évoluent dans le temps, il n'y a que certaines relations entre ces objets, à savoir des relations d'intrication. Ces relations se manifestent sous la forme des fameuses corrélations d'Einstein-Podolsky-Rosen entre des résultats de mesure. Le théorème de Bell établit que les objets quantiques ne peuvent pas avoir de propriétés intrinsèques, même pas des propriétés intrinsèques cachées (des variables cachées), qui fixent les relations d'intrications : s'ils possédaient de telles propriétés intrinsèques, il ne pourrait pas y avoir les corrélations que prédit la théorie quantique et qui sont confirmées par l'expérience⁵.

En ce qui concerne la théorie de la relativité générale, celle-ci ne permet plus de faire une distinction entre matière, d'une part, et espace-temps, de l'autre (où l'espace-temps serait une sorte d'arène passive dans laquelle la matière serait insérée). L'espace-temps, plus précisément le champ métrique, contient lui-même de l'énergie, à savoir l'énergie gravitationnelle. On peut dès lors considérer l'espace-temps – et les points de l'espace-temps entre lesquels s'étendent les relations métriques – comme des objets physiques. Néanmoins, il y a un argument célèbre remontant à Einstein, connu sous le nom d'«argument du trou», qui montre qu'on ne peut pas considérer les points de l'espace-temps comme possédant des propriétés intrinsèques : si les points de l'espace-temps possédaient une identité intrinsèque, indépendamment des relations métriques, alors le déterminisme qu'implique la théorie de la relativité générale ne pourrait pas avoir lieu⁶.

Les deux grandes théories physiques fondamentales du xx^e siècle parlent ainsi en faveur d'une position en philosophie de la nature qui est connue sous le nom de «réalisme structural» : ce qui existe dans le domaine physique fondamental, ce sont en premier lieu des structures, dans le sens de filets de relations physiques concrètes (relations d'intrication quantique, relations métriques) (cf. FRENCH & LADYMAN 2003). Il n'est pas nécessaire de refuser la reconnaissance d'objets dans le domaine physique fondamental (objets quantiques, points de l'espace-temps). Toutefois, les relations qui les relient sont la manière dont ces

5. Cf. pour une introduction à la philosophie de la physique quantique ESFELD (2006, chap. 15 à 18).

6. Cf. ESFELD (2006, chap. 14), pour une introduction.

objets existent, et rien n'empêche que l'existence des objets physiques fondamentaux s'achève dans les relations qui les unissent (cf. ESFELD 2004, ESFELD & LAM 2008).

La mise en évidence de structures, par contraste avec des propriétés intrinsèques, implique les deux conséquences suivantes : (1) L'argument en faveur d'une conception causale des propriétés présenté dans la section précédente ne s'applique pas aux structures. Il n'y a pas de propriétés intrinsèques qui sont des qualités pures sous-jacentes aux structures et qu'on ne peut, par principe, pas connaître. (2) Il est douteux que les structures puissent être causales. Ce sont des structures – et non des propriétés intrinsèques – qu'envisage Russell lorsqu'il soutient que la physique contemporaine bannit la notion de propriétés causales.

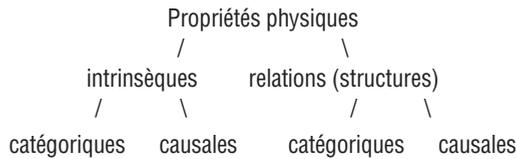
Néanmoins, même si l'on accepte que les théories physiques fondamentales du xx^e siècle se focalisent sur des structures au lieu de le faire sur des propriétés intrinsèques, la situation n'est pas aussi claire que la présente Russell. Rien en physique n'empêche que les structures elles-mêmes soient causales⁷. Autrement dit, dans la mesure où il y a certaines structures physiques qualitatives, celles-ci sont des pouvoirs de produire certains effets. La caractérisation des structures d'intrication quantique en termes dispositionnels est très répandue. Dans la mesure où les structures d'intrication quantique sont des structures physiques qualitatives, elles sont le pouvoir d'engendrer des réductions d'état et, par ce biais, d'engendrer des propriétés physiques classiques (cf. DORATO 2006 et DORATO & ESFELD 2010).

En ce qui concerne les structures métriques, spatio-temporelles, la théorie de la relativité générale donne le coup de grâce à un dualisme insatisfaisant qui postule, d'une part, des propriétés matérielles causales et, d'autre part, des relations spatio-temporelles ne pouvant pas être causales. Comme les structures métriques contiennent de l'énergie, à savoir l'énergie gravitationnelle, on peut les considérer comme causales au même titre que les structures quantiques d'énergie-matière non gravitationnelle : la manière dont le champ métrique interagit avec lui-même ainsi qu'avec l'énergie-matière non gravitationnelle, y compris par la production des phénomènes gravitationnels observés, est la

7. Cf. FRENCH (2006, 178-182), cf. également LADYMAN & ROSS (2007, chap. 2 à 5), qui parlent en termes de structures modales.

manifestation de son caractère causal (cf. BARTELS 1996, p. 37-38, BIRD 2009, section 2.3).

Il faut donc tenir compte de la distinction entre propriétés catégoriques et propriétés causales ainsi que de celle entre propriétés intrinsèques et relations (structures). Toutes les combinaisons sont possibles de sorte que l'espace logique des positions possibles se définit par les quatre positions suivantes :



Les théories physiques contemporaines infirment les deux positions à gauche de ce schéma, mettant l'accent sur des relations (des structures) au lieu de propriétés intrinsèques. Toutefois, en tant que telles, elles laissent ouverte la question de savoir si ces structures sont catégoriques ou si elles sont causales.

Néanmoins, la partie épistémologique de l'argument contre les propriétés intrinsèques catégoriques que nous avons présenté dans la section précédente s'applique dans un certain sens également à la conception russellienne des structures physiques comme catégoriques au lieu de causales. Les structures auxquelles les théories physiques fondamentales font référence ne sont pas observables. Nous observons en physique quantique certaines corrélations entre des résultats de mesure, mais pas de superpositions de telles corrélations, c'est-à-dire pas d'intrications. Nous reconnaissons l'existence des superpositions et des intrications quantiques dans le but d'expliquer les résultats de mesure et les corrélations entre elles. De la même façon, on ne peut pas observer un espace-temps quadridimensionnel courbé. Aucun observateur local ne peut trancher la question de savoir si oui ou non l'espace-temps est courbé. C'est uniquement si l'on met les différentes observations locales ensemble afin de former une représentation de l'espace-temps comme tout qu'on arrive à reconnaître la courbure de l'espace-temps. En bref, les structures physiques fondamentales sont des entités théoriques, et nous les acceptons parce qu'elles expliquent les phénomènes observés. Il s'agit d'une explication causale : les structures physiques fondamentales sont l'origine causale des phénomènes

observés. Ainsi, les structures d'intrication quantique, par le biais des réductions d'état, rendent compte de façon causale des phénomènes observés, les structures métriques rendent compte causalement des phénomènes gravitationnels observés, etc.

Si les structures physiques fondamentales n'étaient pas causales elles-mêmes, c'est-à-dire si elles n'étaient pas des pouvoirs dans le sens expliqué, nous ne pourrions pas les connaître : nous ne pourrions pas savoir quelles sont les structures physiques sous-jacentes. Si les structures physiques fondamentales étaient catégoriques, différentes structures pourraient être corrélées avec les mêmes phénomènes observables. En d'autres termes, sur la base des phénomènes observables, on ne pourrait, par principe, pas savoir quelles sont les structures physiques fondamentales. Il y aurait en ce cas deux mondes possibles distincts quant à leurs structures physiques fondamentales mais indiscernables en ce qui concerne le domaine entier des phénomènes observables dans les deux mondes. Si, par contre, les structures physiques fondamentales sont causales, il n'y a pas une telle sous-détermination des phénomènes observables en entier par les structures physiques fondamentales : chaque différence dans les structures physiques fondamentales est en ce cas une différence causale, et comme les pouvoirs que sont les structures physiques fondamentales s'exercent en engendrant les phénomènes observables, cette différence causale mène forcément à une différence quelque part dans le domaine des phénomènes observables. Nous pouvons donc, par principe, découvrir les structures physiques fondamentales à partir des phénomènes observables.

Cet argument montre que le réalisme structural est un réalisme scientifique si et seulement s'il conçoit les structures physiques fondamentales comme étant des structures causales. Toutefois, en tant que tel, il ne s'agit pas d'un argument en faveur du réalisme scientifique⁸. Néanmoins, on peut, sur la base de cet argument, développer un argument de cohérence de notre savoir comme tout en faveur de la conception causale des structures, voire des propriétés en général.

8. Position selon laquelle les sciences de la nature sont par principe capables de révéler la constitution de la nature.

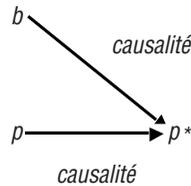
5 ↻ La causalité dans les sciences spéciales

La biologie et la psychologie sont des exemples paradigmatiques de sciences spéciales, par contraste avec les théories physiques fondamentales d'application universelle. Ces sciences considèrent les propriétés dont elles traitent notamment comme des propriétés fonctionnelles, c'est-à-dire comme des propriétés qui sont définies par la production de certains effets caractéristiques (et selon quelques conceptualisations, également par une certaine histoire causale caractéristique). Les gènes, par exemple, se définissent par la production de certains effets phénotypiques, la sélection naturelle concerne certains effets que produisent les propriétés des organismes dans des environnements donnés, etc. De plus, le fonctionnalisme est également la position standard en philosophie de l'esprit : les propriétés mentales se définissent par leur fonction, c'est-à-dire par leurs effets caractéristiques. D'après cette position, le contenu de chaque état intentionnel consiste en des rapports avec d'autres états intentionnels et des actions, et ce qui caractérise les états émotionnels – comme, par exemple, la douleur –, ce sont également leurs effets pour l'organisme (cf. par ex. PUTNAM 1975 [2002]).

La métaphysique humienne peut reconnaître les descriptions fonctionnelles, y compris les théories et les lois, dans lesquelles les concepts fonctionnels des sciences spéciales figurent, comme étant des descriptions vraies du monde, mais elle ne peut pas reconnaître l'existence des propriétés fonctionnelles auxquelles les sciences spéciales font référence. Les propriétés sont catégoriques et, comme expliqué dans la première section de cet article, la distribution des propriétés (ou structures) physiques fondamentales et catégoriques est le vérifacteur des descriptions des sciences spéciales. Autrement dit, ces descriptions ne sont pas vraies parce qu'elles font référence à des propriétés fonctionnelles dans le monde, mais elles sont vraies grâce à certaines régularités contingentes dans la distribution des propriétés physiques fondamentales et catégoriques.

Si l'on recule devant ce matérialisme et ce réductionnisme austères, qui ne reconnaissent en fin de compte que les propriétés physiques fondamentales, on ne peut pas accepter l'existence des propriétés dont traitent les sciences spéciales et les concevoir comme des propriétés émergentes, existant en sus des propriétés (ou structures) physiques fondamentales. En effet, si l'on considère les propriétés dont traitent les sciences spéciales comme étant distinctes des propriétés physiques fon-

damentales, la manière dont celles-ci peuvent figurer dans des relations causales, ayant des effets, n'est pas intelligible. Tout changement dans l'organisme ou son environnement causé par une propriété biologique ou mentale implique un changement physique : tout effet phénotypique que cause un gène implique un changement moléculaire, tout effet que cause une propriété mentale implique un changement moléculaire dans le cerveau, même s'il s'agit d'un effet mental qui n'a pas d'influence directe sur le comportement, etc. Or, pour tout changement moléculaire, qui est ou qui implique à son tour un changement physique au niveau atomique, dans la mesure où ce changement a une cause, il existe une cause physique complète (qu'elle soit exprimée en termes de probabilités objectives ou de façon déterministe est indifférent ici). En bref, si une cause biologique ou mentale b engendre un effet quelconque, il engendre toujours également un effet physique p^* pour lequel il existe aussi une cause physique p :



Or, ce n'est pas une position convaincante que de maintenir que les causes biologiques et mentales surdéterminent systématiquement leurs effets (ces effets possédant toujours également des causes physiques) : toute situation présentant une telle surdétermination régulière serait indiscernable d'une situation dans laquelle les propriétés biologiques et mentales seraient causalement inefficaces, à savoir une situation dans laquelle la flèche de causalité de b à p^* du schéma ci-dessus serait simplement absente ; il y a toujours des causes physiques complètes, et les lois physiques sont des lois plus strictes que les lois biologiques ou psychologiques. En bref, si l'on établit une distinction entre les propriétés biologiques et mentales d'un côté et les propriétés physiques de l'autre, ce sont les propriétés biologiques et mentales qui perdent la compétition causale qui suit de cette distinction⁹.

9. Cf. ESFELD (2005) pour un exposé de ce problème avec discussion des différentes positions.

Si l'on souscrit à la théorie causale des propriétés, on peut reconnaître l'existence des propriétés dont traitent les sciences spéciales et soutenir que ces propriétés sont identiques à des propriétés physiques : la raison en est que les propriétés sont définies de manière causale et qu'il n'y a pas de différence causale entre les occurrences des propriétés des sciences spéciales et les propriétés physiques sous-jacentes. Plus précisément, les structures globales d'intrication quantique se développent par des événements de réduction d'état en des structures locales, et quelques-unes de ces structures locales sont stables et se distinguent de leur environnement parce qu'elles produisent, en tant que tout, des effets pertinents dans un environnement donné. Ces structures locales sont à leur tour des propriétés physiques complexes, et elles sont identiques aux propriétés dont traitent les sciences spéciales car il n'y a pas de différence causale entre les deux. La description physique se focalise sur la composition physique de ces structures, tandis que la description dans le vocabulaire des sciences spéciales se focalise sur les effets pertinents qu'elles produisent comme tout dans un environnement donné. La différence entre ces deux types de description est épistémologique dans le sens où elle n'est qu'une division du travail scientifique, division qui est utile pour l'organisation de la recherche mais qui n'empêche pas l'identité ontologique.

Comme les descriptions fonctionnelles des sciences spéciales sont en règle générale plus abstraites que les descriptions de la composition physique des structures en question, des descriptions des sciences spéciales d'un même type peuvent s'appliquer à des structures physiques de compositions différentes, nécessitant pour cette raison des descriptions physiques différentes. Pourtant, tenant la théorie causale des propriétés comme admise, toute différence de composition physique entre deux structures, nécessitant des descriptions physiques différentes, implique une différence causale dans la façon dont sont produits les effets pertinents de la structure en question. Sur cette base, il est possible de concevoir des concepts fonctionnels des sciences spéciales plus précis qui tiennent compte de ces différences dans la production des effets caractéristiques et ainsi de gagner des concepts fonctionnels des sciences spéciales qui ont la même extension que les concepts physiques décrivant la composition physique des structures en question. De cette manière, il est en principe possible de réduire les théories des sciences spéciales à des théories physiques (cf. ESFELD & SACHSE 2007).

Sur la base de la théorie causale des propriétés, il s'avère dès lors possible de proposer un matérialisme et un réductionnisme non éliminativistes : on reconnaît l'existence des propriétés biologiques, mentales, etc., gardant leurs traits caractéristiques, et on les intègre en même temps dans le monde physique, matériel en soutenant qu'elles sont identiques à des structures physiques locales et que leurs descriptions peuvent être réduites à des descriptions physiques. Même RUSSELL (1912 [2006]), dans sa critique de la conception causale des propriétés, concède que les volitions produisent des effets et que la théorie de la causalité comme production, par contraste avec une théorie humienne de la causalité, s'applique ainsi à la causalité mentale¹⁰. Toutefois, si l'on ne souhaite pas être coincé dans un dualisme désespéré, la seule façon dont une volition peut être causalement efficace est d'être identique à une propriété physique, cérébrale. Cette perspective d'un réductionnisme non éliminativiste qui est en mesure d'intégrer tous les engagements ontologiques de nos théories scientifiques constitue, à notre avis, un argument concluant pour trancher le débat philosophique sur les fondements de la causalité en faveur d'une théorie causale des propriétés, voire des structures, comme existant dans le monde.

☉Références bibliographiques☉

B

BARTELS Andreas (1996), "Modern essentialism and the problem of individuation of spacetime points", *Erkenntnis*, 45: 25-43.

BIRD Alexander (2007), *Nature's metaphysics. Laws and properties*, Oxford University Press.

BIRD Alexander (2009), "Structural properties revisited", in T. Handfield (ed.), *Dispositions and causes*, Oxford University Press: 215-241.

C

CURIEL Erik (2000), "The constraints general relativity places on physicalist accounts of causality", *Theoria*: 33-58.

D

DORATO Mauro (2006), «Dispositions, propriétés relationnelles et monde quantique», in B. Gnassounou & M. Kistler (dir.), *Causes, pouvoirs, dispositions en philosophie. Le retour des vertus dormitives*, CNRS Éditions: chap. 12.

DORATO Mauro & ESFELD Michael (2010), "GRW as an ontology of dispositions", *Studies in History and Philosophy of Modern Physics*, 41B: 41-49.

10. Cf. ESFELD (2007) pour la portée de la causalité mentale quant à la philosophie de la causalité.

DOWE Phil (2000), *Physical causation*, Cambridge University Press.

E

ESFELD Michael (2004), "Quantum entanglement and a metaphysics of relations", *Studies in History and Philosophy of Modern Physics*, 35B: 601-617.

ESFELD Michael (2005), *La Philosophie de l'esprit. De la relation entre l'esprit et la nature*, Armand Colin.

ESFELD Michael (2006), *Philosophie des sciences. Une introduction*, Presses polytechniques et universitaires romandes (2^e éd. revue et mise à jour, 2009).

ESFELD Michael (2007), "Mental causation and the metaphysics of causation", *Erkenntnis*, 67: 207-220.

ESFELD Michael & LAM Vincent (2008), "Moderate structural realism about space-time", *Synthese*, 160: 27-46.

ESFELD Michael & SACHSE Christian (2007), "Theory reduction by means of functional sub-types", *International Studies in the Philosophy of Science*, 21: 1-17.

F

FRENCH Steven (2006), "Structure as a weapon of the realist", *Proceedings of the Aristotelian Society*, 106: 167-185.

FRENCH Steven & LADYMAN James (2003), "Remodelling structural realism: quantum physics and the metaphysics of structure", *Synthese*, 136: 31-56.

G

GRASSHOFF Gerd & MAY Michael (2001), "Causal regularities", in W. Spohn, M. Ledwig & M. Esfeld (eds.), *Current issues in causation*, Mentis: 85-114.

H

HEIL John (2003), *From an ontological point of view*, Oxford University Press.

HUME David (1739), *Traité de la nature humaine*.

HUME David (1748), *Enquête sur l'entendement humain*.

K

KISTLER Max (1999), *La Causalité et les lois de la nature*, Vrin.

KISTLER Max (2005), «L'efficacité causale des propriétés dispositionnelles macroscopiques», in B. Gnassounou & M. Kistler (dir.), *Causes, pouvoir, dispositions en philosophie. Le retour des vertus dormitives*, PUF: 115-154.

L

LADYMAN James, ROSS Don, SPURRETT David & COLLIER John (2007), *Every thing must go. Metaphysics naturalised*, Oxford University Press.

LAM Vincent (2005), "Causation and space-time", *History and Philosophy of the Life Sciences*, 27: 465-478.

LEWIS David (1973a), *Counterfactuals*, Blackwell.

LEWIS David (1973b), "Causation", *Journal of Philosophy*, 70: 556-567. Réimp. in D. Lewis, *Philosophical papers*, vol. 2, Oxford University Press, 1986: 159-172.

LEWIS David (1986a), *Philosophical papers*, vol. 2, Oxford University Press.

LEWIS David (1986b), *On the plurality of worlds*, Blackwell.

LEWIS David (1994), "Lewis, David: Reduction of mind", in S.H. Guttenplan (ed.), *A companion to the philosophy of mind*, Blackwell: 412-431. Réimp. in D. Lewis, *Papers in metaphysics and epistemology*, Cambridge University Press, 1999: 291-324.

LEWIS David (2004), "Causation as influence", in J. Collins, N. Hall & L.A. Paul (eds.), *Causation and counterfactuals*, MIT Press: 75-106.

LEWIS David (2009), "Ramseyan humility", in D. Braddon-Mitchell & R. Nola (eds.), *Conceptual analysis and philosophical naturalism*, MIT Press: 203-222.

LOEWER Barry (2007), "Counterfactuals and the second law", in H. Price & R. Corry (eds.), *Causation, physics, and the constitution of reality. Russell's republic revisited*, Oxford University Press: 293-326.

M

MACKIE John L. (1965), "Causes and conditions", *American Philosophical Quarterly*, 2: 245-264.

MACKIE John L. (1974), *The Cement of the universe. A study of causation*, Oxford University Press.

MUMFORD Stephen (1998), *Dispositions*, Oxford University Press.

MUMFORD Stephen (2006), "The ungrounded argument", *Synthese*, 149: 471-489.

P

PRICE Huw & CORRY Richard (eds.), *Causation, physics, and the constitution of reality. Russell's republic revisited*, Oxford University Press.

PUTNAM Hilary (1975), "The nature of mental states", in H. Putnam (ed.), *Mind, language and reality. Philosophical papers*, volume 2, Cambridge University Press: 429-440. 1^{re} publication: "Psychological predicates" in W.H. Capitan & D.D. Merrill (eds.), *Art, mind and religion*, University of Pittsburgh Press, 1967. Trad. fr., J.-M. Roy & D. Boucher, «La nature des états mentaux», in D. Fisette & P. Poirier (dir.), *Philosophie de l'esprit. Psychologie du sens commun et sciences de l'esprit*, Vrin: 269-287.

R

RUSSELL Bertrand (1912), "On the notion of cause", *Proceedings of the Aristotelian Society*, 13: 1-26. Trad. fr., G. Bourgin, revue par M. Kistler & J. Sackur, «Sur la notion de cause», *Philosophie*, 89, 2006: 3-20.

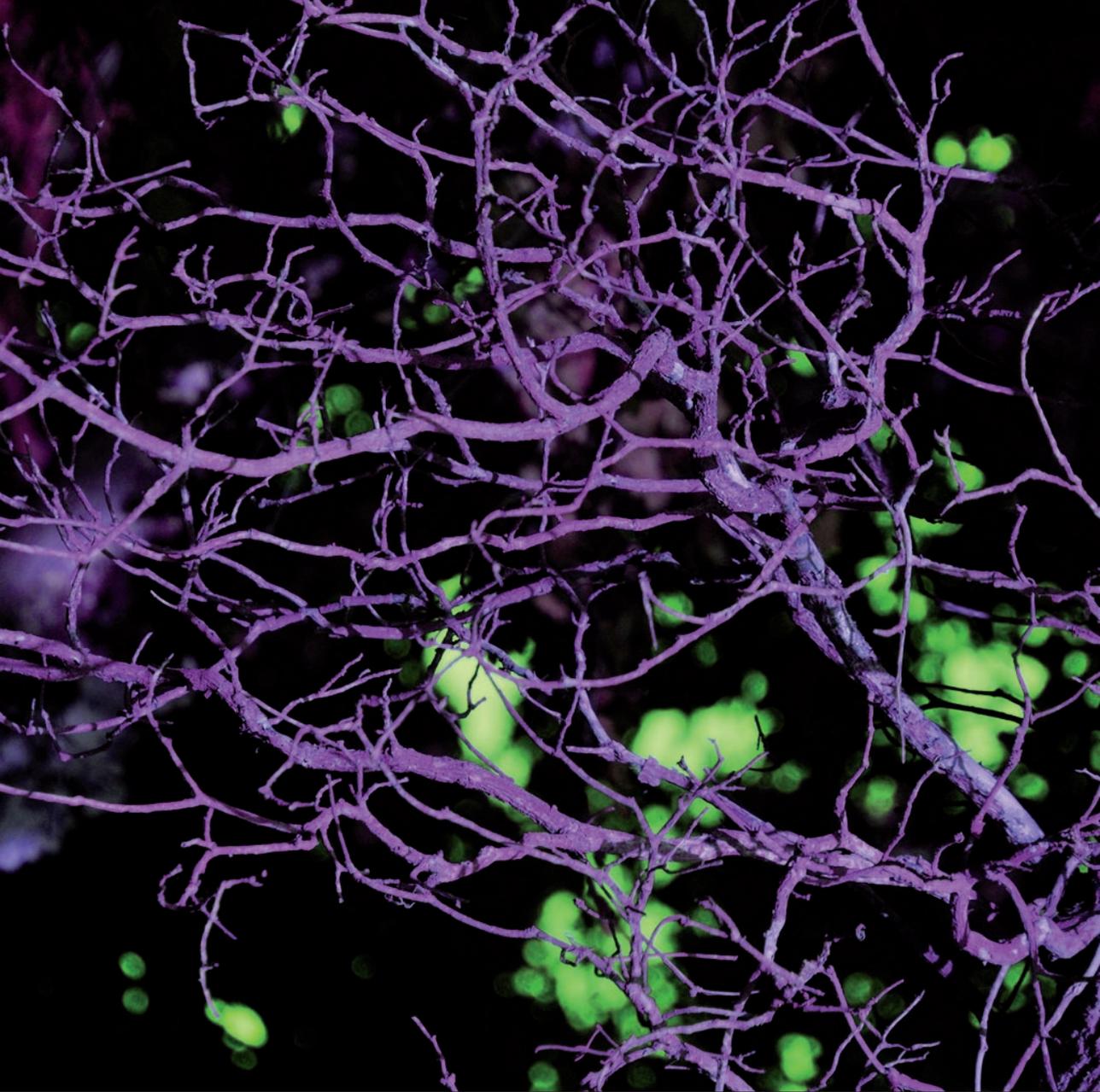
S

SALMON Wesley C. (1998), *Causality and explanation*, Oxford University Press.

SHOEMAKER Sydney (1980), "Causality and properties", in P. van Inwagen (ed.), *Time and cause*, Reidel: 109-135. Réimp. in S. Shoemaker, *Identity, cause, and mind. Philosophical essays*, Cambridge University Press, 1984: 206-233.

☉ Notice biographique ☉

Michael Esfeld est professeur ordinaire de philosophie des sciences à l'Université de Lausanne (Suisse). Ses domaines de recherche sont la philosophie de la physique, la métaphysique des sciences (notamment les thèmes de causalité et de réductionnisme) et la philosophie de l'esprit. Il est membre de Leopoldina, l'Académie nationale de l'Allemagne ; il a reçu le prix Cogito 2008 pour ses travaux visant un rapprochement entre la philosophie et les sciences. Ses publications incluent deux livres en français, *La Philosophie de l'esprit. De la relation entre l'esprit et la nature* (Paris, Armand Colin, 2005) et *Philosophie des sciences. Une introduction* (Lausanne, Presses polytechniques et universitaires romandes, 2006, 2^e éd. 2009).



Avertissement. Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur et des auteurs. Les textes et illustrations figurant sur ce site peuvent être consultés et reproduits sur un support papier ou numérique sous réserve qu'ils soient strictement réservés à un usage personnel, scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document selon ce modèle :

Michael Esfeld, «Les fondements de la causalité», in Gérard Lambert et Marc Silberstein (dir.), Matière première. Revue d'épistémologie [en ligne]. Nouvelle série, N° 1/2010 : Épistémologie de la médecine et de la santé, Éditions Matériologiques. Mis en ligne le 7 octobre 2010.

URL : www.materiologiques.com

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.



Jérôme Ravat **Fonction naturelle et fonction morale : similitude ou imposture ?**

Sujet polémique par excellence, la relation entre valeurs morales et fonctions naturelles représente sans conteste une des questions les plus âprement débattues de la philosophie morale, mais également de la philosophie des sciences.

1 ↻ Le retour du fonctionnalisme moral naturaliste

Longtemps absente au xx^e siècle du débat philosophique, en raison tout particulièrement des critiques dévastatrices du philosophe G.E. Moore, la thématique du naturalisme moral a cependant effectué un retour remarqué au cours des trois dernières décennies, tout particulièrement depuis la publication par l'entomologiste E.O. Wilson du fameux ouvrage *Sociobiology*, en 1975. C'est ainsi que dans la lignée des recherches initiales de la sociobiologie, un nombre pléthorique d'ouvrages et d'articles (tout particulièrement dans le monde anglo-saxon) ont de nouveau abordé la question de la naturalisation de l'éthique. Cette approche naturaliste de l'éthique, à cet égard, comporte plusieurs facettes : sur le plan épistémologique, elle vise à rendre intelligible les phénomènes moraux grâce aux méthodes, et au lexique propres aux sciences naturelles. Sur le plan ontologique, elle a pour but d'identifier les faits moraux à des processus naturels, contre diverses formes de non-naturalisme (tel que le platonisme moral) ou de surnaturalisme (comme les théories du « commandement divin », faisant dépendre les propriétés morales d'un fondement théologique). Il s'agit en ce sens de mettre

notamment en lumière les facteurs biologiques ou psychologiques qui sous-tendent jugements et comportements moraux.

Toutefois, pour certains partisans du naturalisme moral, cette approche scientifique de l'éthique n'a pas uniquement pour vocation de décrire les modalités selon lesquelles se développent les phénomènes moraux. Plus fondamentalement, elle pourrait permettre (du moins partiellement) de fonder et de justifier l'adoption de certaines normes, censées être davantage en concordance avec la manière dont la nature humaine se manifeste et s'épanouit : la naturalisation de l'éthique, à ce titre, aurait pour vocation de définir les limites naturelles de nos systèmes normatifs, ou de résoudre un certain nombre de désaccords moraux à partir d'une analyse naturaliste des phénomènes moraux. C'est ce versant prescriptif, bien évidemment, qui suscite les polémiques les plus virulentes à l'encontre du naturalisme¹.

À cet égard, le concept de fonction, transposé du domaine biologique vers le domaine moral, constitue une des pièces maîtresses dans la stratégie d'élaboration d'une éthique normative à partir de théories naturalistes. C'est ainsi qu'un certain nombre d'auteurs, s'efforçant résolument de remettre en cause la critique mooréenne du sophisme naturaliste (cf. section 2), se sont évertués à développer l'idée selon laquelle les fonctions naturelles caractérisant l'espèce humaine pouvaient permettre de rendre intelligible un certain nombre de nos fonctions morales : autrement dit, pour les partisans de ce que nous pourrions nommer le «fonctionnalisme moral naturaliste», l'évolution biologique constitue le processus naturel à partir duquel il devient possible d'envisager les obligations, les normes et les comportements moraux les plus en harmonie avec la nature morale de l'espèce humaine.

1. En guise d'exemple de ces polémiques, il suffit de se rappeler que les thèses des sociobiologistes furent utilisées au cours des années 1970 en France par les tenants de ce que l'on a nommé la «Nouvelle droite». Ce groupuscule d'intellectuels, farouchement hostile à l'État républicain, prétendait notamment que les découvertes effectuées par les sociobiologistes permettaient de justifier les hiérarchies sociales et les inégalités s'y rattachant. La Nouvelle droite, à cet égard, fut souvent accusée de légitimer les inégalités raciales. Il faut toutefois préciser que la récupération de la sociobiologie par la Nouvelle droite repose sur une série d'interprétations très contestables des recherches initiales de E.O. Wilson et de ses collaborateurs.

C'est cette nouvelle forme de naturalisme moral, faisant un usage prescriptif de la notion de fonction biologique, que nous nous proposons ici d'examiner et de critiquer, en centrant notre analyse sur un ouvrage récent au titre révélateur, *Natural Ethical Facts*, de William CASEBEER (2005). Plus précisément, il s'agira pour nous, par le biais d'une critique de la notion de fonction naturelle, de mettre l'accent sur le fait que la référence à l'évolution biologique s'avère en de nombreux points inadéquate pour fonder de manière absolue normes et valeurs morales. Le fonctionnalisme moral naturaliste, comme nous le verrons, commet en effet trop de confusions conceptuelles, de parti pris idéologiques et de généralisations hâtives pour pouvoir être soutenu de manière pleinement convaincante et permettre l'édification d'une éthique normative.

2 ⇨ Critique mooréenne du naturalisme moral et réponse du réalisme moral naturaliste

Avant de nous intéresser à la théorie de Casebeer, il importe de souligner que toutes les tentatives effectuées depuis le début du xx^e siècle pour rattacher valeurs morales et nature humaine se sont heurtées (le plus souvent de manière fatale) au très fameux argument du «sophisme naturaliste», développé par le philosophe G.E. Moore dans *Principia Ethica* (1903). Comme l'explique Moore dans cet ouvrage, il est parfaitement illégitime de réduire des faits moraux à un ensemble de faits naturels car le bien moral («good») ne peut être identifié à une propriété naturelle particulière. Ainsi, il est impossible de déterminer si des phénomènes naturels comme le plaisir, le bonheur ou la complexité biologique peuvent être identifiés au bien. Comme l'explique MOORE (1903, chap. 13), l'indétermination du bien conduit à formuler à son égard une «question ouverte»: lorsque nous nous demandons par exemple «Est-ce que X est bien?», il est toujours possible de remplacer «X» par une propriété descriptive quelconque (comme le plaisir ou la complexité biologique) sans que cette question ne perde son sens. Le sens d'une telle question n'est donc pas prédéterminé *a priori*, de sorte qu'il est impossible par la seule analyse conceptuelle d'identifier le bien à une propriété naturelle déterminée.

Il est important ici de rappeler que la critique formulée par Moore visait un certain nombre de doctrines philosophiques (comme l'utilitarisme hédoniste de Bentham) mais aussi l'éthique évolutionniste

de Herbert Spencer, qui rattachait adaptation fonctionnelle et valeur morale. Comme l'expliquait en effet Spencer dans *The Data of Ethics* (1879), supériorité morale et complexité fonctionnelle sont des notions tout à fait synonymes car il existe un parallélisme étroit entre complexité biologique et aptitude morales : à la dynamique de complexité adaptative et fonctionnelle qui caractérise les organismes vivants au cours de l'évolution, répond une logique de moralisation progressive, en vertu de la «loi de l'évolution». De sorte que pour Spencer, plus un organisme est fonctionnellement adapté, plus il est moralement évolué, sa capacité adaptative s'avérant étroitement solidaire de ses aptitudes morales. Or, insiste Moore, le fonctionnalisme moral de Spencer ne saurait être défendu, car il procède de manière illégitime à une identification du bien moral à un ensemble de fonctions biologiques. Une telle identification s'avère selon Moore hautement problématique, car en vertu de l'argument de la question ouverte, le bien moral pourrait parfaitement ne pas se réduire à la complexité adaptative et fonctionnelle sans que les énoncés le décrivant ne perdent pour autant leur signification. Selon Moore, on ne saurait donc assimiler bien moral et fonction naturelle.

On aurait pu croire qu'à l'issue de la critique de Moore, le projet d'enraciner les normes morales dans l'évolution biologique serait purement et simplement abandonné. Pourtant, il n'en est rien. Prenant appui sur de nouvelles théories en philosophie des sciences et en méta-éthique² certains auteurs, défendant un réalisme moral naturaliste, se sont efforcés de répondre aux objections initiales de Moore. Ainsi, selon les partisans de ce que l'on appelle le «réalisme moral naturaliste» ou «réalisme de Cornell» (Richard Boyd, Peter Railton, David Brink), il est parfaitement possible de rattacher les valeurs morales (en particulier le bien) à un ensemble de propriétés naturelles. Cette identification, affirment-ils, ne saurait certes être établie de manière conceptuelle (en ce sens, les réalistes moraux naturalistes tiennent compte des critiques de Moore), mais elle peut toutefois s'effectuer de manière empirique, grâce aux découvertes effectuées par le biais des sciences de la nature. Une analogie simple, souvent utilisée par les réalistes moraux natura-

2. La méta-éthique est un domaine de la philosophie morale qui s'intéresse à l'origine et à la signification des concepts moraux. La réflexion méta-éthique porte par exemple sur la signification de concepts comme «bon», «juste», «devoir», etc.

listes, et empruntée au domaine de la physique, permet de comprendre ce point : le fait que l'eau soit composée d'un atome d'oxygène et de deux atomes d'hydrogène n'est pas une question définitionnelle, mais empirique. Autrement dit, l'identité entre le concept « eau » et le phénomène chimique « H₂O » n'est pas établie *a priori*, par le seul travail définitionnel : elle s'effectue *a posteriori*, grâce à des découvertes empiriques qui ont permis par exemple de passer de la conception aristotélicienne (finaliste) au paradigme physico-chimique contemporain.

De manière similaire, selon les tenants du réalisme moral naturaliste, le prédicat « bon » pourrait être identifié, grâce à un certain nombre de découvertes empiriques (par exemple dans le domaine biologique ou psychologique) à un ensemble de propriétés naturelles. Cette découverte, du reste, n'a pas seulement des implications méta-éthiques. Elle pourrait avoir des conséquences normatives, par exemple en permettant de résoudre les conflits moraux au moyen d'une investigation naturaliste. Comme l'écrit à ce propos Richard Boyd, de manière très optimiste :

Un examen philosophique attentif indiquera que des accords sur des faits non moraux élimineront presque tous les désaccords moraux concernant les questions morales qui apparaissent dans les pratiques morales ordinaires. (BOYD 1988, p. 213, notre traduction.)

3 ↻ Le fonctionnalisme moral de William Casebeer

C'est donc dans la droite lignée des défenseurs du réalisme moral naturaliste que Casebeer s'efforce de développer une synthèse originale entre l'éthique de la vertu d'Aristote et les acquis récents de la biologie néodarwinienne. Reprenant explicitement le projet du réalisme de Cornell, Casebeer affirme en effet que les découvertes scientifiques pourraient permettre de découvrir un certain nombre de faits moraux objectifs :

De même que l'on peut développer une connaissance médicale, il est possible de développer une connaissance morale ; cette dernière sera acquise de la même manière que la connaissance scientifique, en raisonnant sur les données de l'expérience. (CASEBEER 2005, p. 48, notre traduction.)

À cet égard, la notion de fonction (« *ergon* »), telle qu'elle est employée dans la philosophie grecque, occupe une place centrale dans ce projet de compréhension naturaliste de l'éthique. Pour Casebeer, les faits moraux

sont identifiables à un ensemble de faits naturels, et en particulier à des phénomènes d'ordre fonctionnel; ils ne sont donc nullement irréductibles aux faits étudiés par les sciences naturelles et ne représentent pas, à l'inverse de ce qu'affirme par exemple John Mackie, des propriétés ontologiquement mystérieuses³.

Les faits moraux sont des faits fonctionnels et les faits fonctionnels ne sont pas étranges; nous pouvons parfaitement les comprendre dans le cadre d'une ontologie matérialiste. (*Ibid.*, p. 55, notre traduction.)

Par ailleurs, comme le précise Casebeer, cette naturalisation des phénomènes moraux est parfaitement compatible avec le réalisme moral, même s'il ne s'agit pas bien évidemment pas d'un réalisme moral de type platonicien, mooréen, ou d'ordre théologique :

Nous pouvons en ayant bonne conscience être des réalistes moraux et pourtant accepter de manière humble notre manière de connaître le bien; une telle humilité résulte non seulement des contraintes pesant sur notre architecture cognitive mais aussi des contraintes qui pèsent sur la forme des processus évolutifs permettent de fixer naturellement les normes. (*Ibid.*, p. 4-5, notre traduction.)

Afin de comprendre ce point, il importe en premier lieu de rappeler brièvement le sens du concept de « fonction » dans la pensée aristotélicienne; dans la cosmologie finaliste d'Aristote, en effet, chaque objet naturel remplit une fonction spécifique. Ainsi, la fonction de l'œil est-elle de bien voir, celle du cheval de courir, celle du couteau de trancher, etc. De même, pour Aristote, la fonction de l'homme est de vivre conformément à la raison, et de développer un certain nombre de vertus morales et intellectuelles comme le courage, l'amitié ou la générosité. C'est avant tout dans la sphère politique, selon Aristote, que la fonction (*l'ergon*) propre à l'homme est la mieux à même de s'exercer. C'est en tant qu'animal politique, en d'autres termes, que l'homme est le plus à même de favoriser l'épanouissement de ses aptitudes naturelles.

3. Dans *Ethics: Inventing Right and Wrong* (1977), John Mackie niait l'existence de faits moraux objectifs en affirmant que de tels faits seraient « bizarres » (*weird*) : selon lui, en effet, de tels faits seraient à la fois descriptifs et prescriptifs, ce qui rend leur statut ontologique extrêmement mystérieux. En identifiant faits moraux et faits fonctionnels, Casebeer tente de répondre à l'objection de Mackie dans le cadre du naturalisme évolutionniste.

Or, selon Casebeer, la théorie de l'évolution permet de reconsidérer la question des fonctions proprement humaines, sans adhérer pour autant aux conceptions téléologiques propres à la pensée aristotélicienne. Ainsi, dans le cadre de la théorie de l'évolution, il devient possible de comprendre, sans recours à un quelconque discours finaliste, la fonction des êtres naturels, et ce en mettant l'accent sur les causes de leur existence au regard de la sélection naturelle.

Afin d'illustrer cette idée, Casebeer mentionne deux conceptions fondamentales de la notion de fonction biologique, la conception causale et la conception étiologique⁴. Selon la version standard de la théorie étiologique, développée notamment par Larry WRIGHT (1976) et par Ruth MILLIKAN⁵ (1984), Z a pour fonction de faire F si le fait que les précédents Z ont eu l'effet F a contribué à leur existence, à leur maintien et à leur diffusion, à partir d'un certain nombre de mécanismes, notamment la sélection naturelle. Ainsi, au moyen de la théorie néodarwinienne, il est possible de comprendre que la fonction d'un organe comme le cœur est de faire circuler le sang, car c'est cette fonction qui est la cause de l'existence et du maintien de cet organe. Conformément à la théorie étiologique, un cœur malade a bien pour fonction de faire circuler le sang, et ce même s'il ne le fait pas effectivement circuler. De même, dans cette perspective, les pouces opposés ont pour fonction de saisir des objets, les motifs tachetés sur les ailes des phalènes ont pour fonction de les camoufler, les magnétosomes (granules d'oxydes métalliques) des bactéries magnétoctatiques ont pour fonction l'orientation dans le champ magnétique terrestre ; dans tous les cas, la fonction d'une entité biologique se trouve référée à des conditions d'effectuation ayant eu lieu au cours de l'évolution.

Selon la conception causale de la fonction, développée notamment par Robert CUMMINS (1975), la fonction d'un objet naturel ou artificiel peut être définie au regard de son rôle présent, c'est-à-dire de sa capa-

4. Le débat sur les fonctions biologiques, depuis les années 1970, est très riche et assez complexe. Dans le cadre de cet article, il ne s'agit nullement pour nous de proposer une analyse exhaustive de ce débat, mais plus modestement d'expliquer de quelle manière il a été transposé par certains auteurs (comme Casebeer) dans le champ de la philosophie morale.

5. Cf. l'analyse que fait Nicolas SHEA (2006) de «la contribution de Ruth Millikan à la philosophie matérialiste de l'esprit». (NdÉ.)

cité, au sein d'un système à produire un certain nombre d'effets. Par exemple, la fonction d'une paire de ciseaux est de couper, et c'est pour cette raison que les ciseaux existent. Selon la conception causale de la fonction, il n'est donc pas nécessaire de retracer le parcours phylogénétique d'un organe ou d'un être vivant pour en développer une analyse fonctionnelle : fonction et évolution sont ici dissociées.

Toutefois, selon Casebeer, l'analyse étiologique et l'analyse causale soulèvent toutes deux d'importantes difficultés : l'analyse étiologique, en rattachant la fonction d'un être à sa phylogenèse risque d'imputer des fonctions qui ne sont plus valables ou utiles dans l'environnement présent. L'analyse causale, quant à elle, commet l'erreur inverse en ne tenant pas suffisamment compte des dynamiques temporelles qui ont conduit à l'émergence de certaines fonctions. Selon Casebeer, c'est en ayant recours à une théorie mixte, insistant de manière modérée sur l'importance des processus phylogénétiques, qu'il devient possible de comprendre la fonction proprement humaine. Cette conception mixte, précise Casebeer, correspond à la théorie développée par Godfrey-Smith, selon laquelle la fonction correspond à un certain nombre d'effets produits au cours d'une histoire évolutive récente. Comme l'explique en effet Godfrey-Smith, cité par Casebeer, les fonctions biologiques sont « [...] des dispositions et des capacités qui expliquent le maintien récent d'un trait dans un contexte sélectif » (CASEBEER 2005, p. 53, notre traduction).

Prenant appui sur cette conception de la fonction, Casebeer en déduit que c'est en se penchant sur les facteurs qui ont permis l'évolution récente d'*Homo sapiens* qu'il est possible de comprendre la fonction proprement humaine. Par là même, il devient possible selon Casebeer d'éviter les difficultés inhérentes à la théorie causale et à la théorie étiologique.

S'il s'avère donc possible, au moyen de la théorie darwinienne, de comprendre la fonction d'un organe comme le cœur, ou de comprendre la fonction d'organismes biologiques plus complexes, qu'en est-il s'agissant des fonctions proprement humaines, en particulier sur le plan moral ? Prenant appui sur l'éthique de la vertu d'Aristote, Casebeer affirme que la fonction propre à l'espèce humaine n'est autre que la socialité, clef de voûte de l'épanouissement personnel : par exemple, le fait de multiplier les relations amicales ayant pour but le développement de l'excellence éthique (ce qu'Aristote, au livre VIII de l'*Éthique à Nicomaque*, nomme les « amitiés vertueuses ») constitue selon Casebeer un moyen fonda-

mental pour l'épanouissement des facultés morales et des traits de caractère qui s'y rattachent : c'est en développant les amitiés vertueuses, par conséquent, que l'espèce humaine a pu, selon Casebeer, survivre et se développer dans un contexte sélectif récent. En outre, précise-t-il, c'est également le développement des interactions sociales qui permet avant tout à la connaissance morale de s'accroître, notamment au moyen d'exemples de comportements vertueux. En effet, c'est par l'habitude suffisamment exercée que les « vertus de caractère », comme le courage, sont produites par le bais d'interactions sociales et par la confrontation avec des exemples de comportements vertueux. C'est donc en développant des traits de caractère vertueux, et en adoptant un certain nombre de fonctions sociales propres à épanouir son statut d'« animal politique » que l'homme a pu survivre et s'épanouir dans son environnement récent. De sorte que pour Casebeer, c'est avant tout grâce au développement des liens sociaux que l'espèce humaine est capable de remplir la fonction qui lui a été en quelque sorte « léguée » par l'évolution.

En bref, le fonctionnalisme évolutionniste permet bel et bien de fonder l'objectivité morale dans la mesure où les faits moraux qu'il décrit ne relèvent pas d'une quelconque réalité transcendante mais peuvent être appréhendés uniquement grâce aux ressources du naturalisme et aux apports de la théorie néodarwinienne : l'évolution de l'espèce humaine a permis l'émergence d'une fonction proprement humaine, et cette fonction est le résultat d'un processus évolutif aléatoire, et non d'une finalité inhérente aux processus naturels. Cette fonction proprement humaine, conformément à la théorie aristotélicienne, réside dans la capacité à utiliser la rationalité pratique au service de relations sociales harmonieuses.

Cette identité entre faits moraux et fonctions naturelles, nous le voyons, a des conséquences sur le plan normatif : elle peut permettre, conformément au projet du réalisme de Cornell, d'apporter des solutions aux désaccords moraux. La naturalisation de l'éthique au moyen du concept de fonction pourrait permettre de « [...] cerner et de résoudre des conflits moraux et de nous permettre d'acquérir des méthodes pour développer nos connaissances morales en renforçant l'efficacité de nos institutions collectives » (CASEBEER 2005, p. 3, notre traduction).

Au moyen d'une approche fonctionnelle de la nature humaine, basée sur les acquis du darwinisme, il serait donc possible, affirme Casebeer, de circonscrire le champ d'application légitime de nos actions. Cette

approche fonctionnaliste des phénomènes moraux, toutefois, est-elle valide? Fonction morale et fonctions naturelles sont-elles interdépendantes, comme l'affirme Casebeer? Selon nous, la position de Casebeer recèle trop de difficultés et de confusions conceptuelles pour pouvoir être légitimement soutenue, et ce pour de multiples raisons.

4 ↻ Limites du parallélisme entre fonction biologique et fonction morale

Une première difficulté du rapprochement entre fonctions naturelles et valeurs morales (et non des moindres) réside dans la valeur même de l'analogie entre le domaine biologique et le domaine éthique.

Afin de se rendre attentif à ce point, il suffit de considérer le cas d'une fonction biologique comme celle consistant à faire circuler le sang pour le cœur. S'il est possible d'affirmer que la fonction biologique du cœur n'est autre que de faire circuler le sang (dans la mesure où cette activité constitue une cause, voire la cause de l'existence de cet organe), doit-on pour autant en déduire que cette fonction est bonne d'un point de vue moral? Imaginons par exemple que pour une raison quelconque un cœur ne soit plus en mesure de pomper le sang. Ce dysfonctionnement biologique constitue-t-il une transgression morale quelconque? La biologie évolutionniste peut certes nous permettre de comprendre en quoi un cœur qui fait circuler le sang fonctionne correctement, mais en faisant cela, elle ne nous donne aucune raison d'affirmer qu'il a l'obligation de le faire, ou qu'il commis a commis une faute s'il n'accomplit pas cette fonction! Comme l'écrit à ce propos, et avec une certaine ironie, Richard Joyce dans *The Evolution of Morality*, lorsque nous examinons la fonction biologique du cœur, «[...] nous ne formulons pas plus d'éloge moral que si nous félicitons un tueur pour son efficacité impitoyable en admettant qu'il est très bon dans l'exercice de sa profession» (JOYCE 2006, p. 171, notre traduction).

Le même raisonnement peut être tenu pour les fonctions biologiques propres aux organismes humains. Nul ne saurait douter par exemple que se reproduire constitue une des fonctions biologiques d'*Homo sapiens*. Pour autant, faudrait-il prononcer une condamnation morale à l'égard de ceux qui ne décident pas d'accomplir cette fonction biologique? Le risque, alors, serait de faire usage du discours fonctionnaliste (dont la légitimité biologique est incontestable) afin de stigmatiser moralement un certain nombre de comportements jugés «déviant». Or, ce passage

de la fonction naturelle à la fonction morale ne va nullement de soi. Par exemple, s'il est possible d'affirmer, sans susciter un tollé, qu'une des fonctions biologiques de la femme est de maximiser son potentiel génétique en donnant naissance à une progéniture, il semble beaucoup plus délicat d'affirmer qu'il s'agit là d'une fonction morale et que toute femme a le devoir d'enfanter ! La chose pourrait même prêter à rire si nombre d'organisations religieuses fondamentalistes (notamment celles qui entreprennent de combiner créationnisme et théorie de l'évolution⁶) n'utilisaient pas précisément cet argument pour stigmatiser la sexualité ne visant pas la reproduction, et les relations pratiquées hors mariage. Faudrait-il, en suivant un tel principe, condamner les individus qui décident en toute connaissance de cause de ne pas perpétuer l'espèce humaine, en s'adonnant à la contraception ? Faudrait-il de même condamner les homosexuels sous prétexte qu'ils ne contribuent pas à la perpétuation de l'espèce et qu'ils commettent ce faisant la transgression d'une norme instituée par la nature ? S'agissant des homosexuels, une telle condamnation pourrait encore une fois prêter à rire si elle n'avait pas été explicitement formulée à plusieurs reprises dans les débats publics récents. Souvenons nous par exemple des propos du député Christian Vanneste (UMP), affirmant dans le journal *La Voix du Nord* en 2005 que l'homosexualité constituait une « menace pour la survie de humanité » et représentait (parce qu'elle ne permet pas la reproduction) une forme de sexualité « inférieure » à l'hétérosexualité. La menace pour l'humanité, si elle existe, résiderait plutôt ici dans la bêtise de tels propos homophobes, dont l'ineptie n'a d'égale que la dangerosité... Un dernier cas, tout aussi intéressant (mais sans doute plus polémique) peut ici aussi être évoqué : la réprobation de l'inceste. Ici encore, la fonction biologique d'une telle réprobation peut être aisément comprise : la consanguinité soulève en effet de nombreux problèmes sur le plan biologique, et on peut comprendre qu'un mécanisme d'aversion, renforcé par une émotion comme le dégoût, ait pu se développer au cours de l'évolution biologique, comme l'affirment certains auteurs⁷. Mais si le rejet des pratiques incestueuses revêt sans nul doute une fonction

6. Évolutionnisme théiste, *Intelligent Design*, etc. Cf. la mise au point de BAUDOUIN & BROUSSEAU (2011, à paraître chez Belin). (NdÉ.)

7. Pour un examen intéressant de cette hypothèse, cf. par exemple FAUCHER (2007).

biologique (à savoir empêcher la dégénérescence génétique), cela signifie-t-il pour autant qu'il s'agit d'une fonction morale? Que penser par exemple d'un inceste entre adultes consentants, et qui décideraient de ne pas se reproduire? Ne s'agit-il pas en l'occurrence d'un «crime sans victime», dans lequel personne n'est lésé, et dont la supposée «immoralité» pose problème?

En somme, loin d'aller nécessairement de pair, fonction morale et fonction biologique doivent être impérativement distinguées, afin d'éviter toute confusion conceptuelle ou pratique. Nombre de nos intuitions fondamentales, si «fonctionnelles» ont-elles pu être au cours de notre passé biologique, ne sont pas pour autant utiles ou moralement défendables.

5 ↻ Difficultés du perfectionnisme naturaliste

Aux objections précédemment avancées, les partisans du fonctionnalisme moral pourraient rétorquer que notre constat est excessivement sévère, dans la mesure où la théorie de Casebeer s'intéresse avant tout à ce qui a permis l'épanouissement récent de l'espèce humaine, faisant jouer à cet égard un rôle secondaire aux facteurs strictement biologiques. Mais ici encore, le fonctionnalisme moral naturaliste se heurte à des difficultés d'importance: comment déterminer, en effet, si tel ou tel comportement est le mieux à même de favoriser l'épanouissement de l'essence morale de l'humanité? Casebeer affirme que c'est la socialité qui permet le mieux à notre espèce de réaliser sa fonction, mais pour quelles raisons cette caractéristique serait-elle la plus fondamentale au regard de l'épanouissement de l'espèce humaine? Il s'agit, ici encore, d'une «question ouverte», dont la réponse s'avère loin d'être frappée du sceau de l'évidence. Autrement dit, le fonctionnalisme de Casebeer soulève la même difficulté que nombre de théories naturalistes cherchant à déterminer l'essence morale de l'homme: elle court le risque de l'arbitraire. Tous les partisans de l'éthique de la vertu se heurtent à cet obstacle majeur, et Casebeer bien évidemment ne saurait échapper à cette règle.

Pour se rendre attentif à ce point, il suffit de comparer la théorie de Casebeer avec d'autres théories s'efforçant également de déterminer objectivement, au moyen des acquis de la biologie évolutionniste, la nature morale de l'espèce humaine. Tel est le cas par exemple de la position de Larry Arnhart (professeur de sciences politiques, proche des

milieux conservateurs américains), exposée dans *Darwinian Natural Right* (1998). Selon lui, l'évolution biologique a doté l'espèce humaine d'un certain nombre de désirs fondamentaux (comme le désir de prendre soin de sa progéniture, le désir de justice, le désir d'éprouver du plaisir esthétique, le désir de religiosité, mais aussi le désir de faire la guerre). La nature humaine ne peut s'épanouir pleinement que si elle est en mesure de développer ces désirs. Même si elle peut rejoindre la théorie de Casebeer sur certains points, la perspective de Arnhart s'en distingue également à différents égards : par exemple, l'assouvissement d'un désir fondamental comme celui de faire la guerre ne s'accorde pas nécessairement avec le développement des amitiés vertueuses. Le désir de religiosité peut être satisfait indépendamment de toute interaction sociale, par exemple dans l'isolement méditatif. Comment décider, à l'aune de ces divergences flagrantes, qu'une conception naturaliste est plus acceptable qu'une autre, plus à même de circonscrire la fonction morale de l'homme ? La définition de l'essence morale de l'humanité semble ici profondément entachée de subjectivité, variant considérablement selon les engagements ontologiques et normatifs auxquels souscrivent les auteurs.

De toute façon, à supposer que l'on soit capable, comme le souhaite Casebeer, de spécifier le domaine des vertus fondamentales à partir d'une analyse biologique de la notion de fonction (tâche sinon impossible, du moins extrêmement délicate), cela donnerait-il pour autant l'obligation morale de se conformer aux phénomènes naturels ainsi découvertes ? Faudrait-il condamner par exemple un individu qui ne chercherait pas à développer ses vertus, sous prétexte qu'il ne réaliserait pas ce faisant l'essence morale de l'humanité en s'abstenant d'accomplir ce qui a permis à ses ancêtres de survivre dans un passé récent ? Comme le précise clairement JOYCE (2006, p. 172, notre traduction) :

Même si les données paléanthropologiques démontraient que les pressions sélectives majeures ayant permis l'émergence du sens moral s'identifiaient au besoin de regrouper des hominidés afin qu'ils puissent s'engager dans des relations de coopération les uns avec les autres [...], peu d'entre nous en concluraient que nous devrions maintenant respecter et privilégier cette fonction.

Ici encore, nous le voyons, normativité morale et normativité biologique semblent se heurter plutôt que d'être en parfaite harmonie.

6 ⇨ Valeur morale et fonction biologique à la lumière des dynamiques de l'évolution : diagnostic d'un profond hiatus

Au fond, loin de justifier une morale ancrée dans l'évolution biologique, le fait de se pencher sur les soubassements naturels des phénomènes moraux permet au contraire de prendre partie en de nombreux points contre le naturalisme moral, tant le hiatus entre processus naturels et moralité paraît important à maints égards. Si une analyse pertinente du fonctionnalisme évolutionniste permet en ce sens de comprendre en quoi certaines tendances comportementales ont pu revêtir une valeur adaptative à un moment donné de notre phylogenèse, c'est également pour mettre en relief le fossé entre cette valeur adaptative passée et nos valeurs morales actuelles.

De surcroît, toute analyse de l'articulation entre valeur morale et fonction biologique doit impérativement tenir compte de la dimension profondément contingente du processus évolutif, à la différence de la lecture téléologique chère à Spencer. Les apports de la biologie évolutionniste, induisant le rejet d'un quelconque sens de l'évolution, invalident en effet l'idée spencérienne d'une valeur morale des processus naturels. Comme l'a expliqué par ailleurs Stephen Jay GOULD (2001), il est problématique d'assimiler complexité biologique et adaptation : après tout, les créatures les mieux adaptées à leur environnement ne sont autres que les bactéries ! Bien évidemment, ce n'est pas pour autant que cet état de fait leur octroie une valeur morale supérieure.

Bien plus, dans la mesure où l'émergence des faits moraux s'avère étroitement dépendante du processus éminemment aléatoire qu'est l'évolution biologique, d'autres tendances naturelles ayant des répercussions dans le domaine moral auraient pu s'avérer davantage prégnantes. Ainsi, il est parfaitement possible d'imaginer que l'évolution biologiques ait permis l'émergence de normes sociales favorisant en priorité les comportements belliqueux ou xénophobes. De telles normes auraient alors été fonctionnelles, mais auraient-elles pour autant été morales ? Ce problème, du reste, avait été déjà perçu par Darwin lui-même, qui affirmait que l'espèce humaine aurait pu évoluer comme celles des abeilles et comporter une tolérance pour le fratricide. Dans un tel cas, la fonctionnalité biologique du fratricide aurait-elle entraîné *de facto* sa valeur morale ? Affirmer la profonde contingence des trajectoires empruntées par l'évolution biologique conduit donc de manière inévitable à dissocier fonctions naturelles et valeurs morales.

Certains partisans d'un fonctionnalisme moral naturaliste pourraient alors rétorquer qu'au moins certaines tendances comportementales issues de l'évolution pourraient revêtir à la fois une fonction morale et une fonction biologique. Un exemple typiquement envisagé dans un tel cadre est celui de la tendance humaine à la coopération⁸. Si l'on envisage des interactions individuelles réitérées dans le temps, comme dans le cadre de la théorie des jeux, alors certaines formes de coopération pourraient émerger, comme le fameux «*tit for tat*» (stratégie consistant à coopérer avec les individus qui ont coopéré et à ne plus coopérer si ces individus font défection). La coopération, à cet égard, est une «stratégie évolutionnairement stable» s'avérant par là même hautement fonctionnelle d'un point de vue biologique. C'est ce qu'indiquait la théorie de l'altruisme réciproque défendue par TRIVERS (1971) et reprise par AXELROD (1984).

Pourtant, ici encore, fonction biologique et fonction morale doivent impérativement être distinguées. Certes, on peut affirmer que dans presque tout monde possible, à condition que les individus disposent de capacités cognitives minimales, des formes particulières de coopération vont pouvoir évoluer (SKYRMS 1996) de sorte que certaines tendances morales (comme la valorisation de l'altruisme ou la condamnation des tricheurs égoïstes) vont émerger de manière mécanique. Toutefois, si la coopération et l'altruisme (qu'il soit ou non réciproque) revêtent sans nul doute une fonction biologique (permettant aux individus et aux groupes de survivre et de se pérenniser), sa valeur morale n'est pas aussi indéniable qu'il pourrait sembler, en raison tout particulièrement de l'indétermination du champ d'extension des liens coopératifs. Nous pouvons avoir tendance à coopérer dans des groupes plus ou moins étendus, et qui ne sont pas forcément en harmonie les uns avec les autres (famille, clan ethnique, groupe religieux, nation, etc.). Autrement dit, cette tendance à la coopération peut revêtir une fonction biologique (la survie de l'individu) ou sociale (la pérennité des institutions) sans pour autant revêtir une fonction morale, car la coopération est une notion relative : nous coopérons avec des individus, mais également contre d'autres individus. En période de guerre, par exemple, les liens coopératifs sont fortement renforcés au sein d'une nation (la guerre, selon

8. Sur cet important domaine, cf. l'état de la question dressé par CLAVIEN (2009). (NdÉ.)

l'expression de Hegel est la «santé des peuples»), mais ce n'est que pour mieux combattre une autre nation avec laquelle aucune coopération ne s'établit. Le fait que la coopération revête une fonction biologique ou sociale ne suffit donc pas à lui seul pour assigner à ce comportement une valeur morale.

Enfin, il convient de garder à l'esprit que les travaux en théorie évolutionniste des jeux indiquent également que dans certaines conditions, les comportements coopératifs ou altruistes ne sont pas nécessairement fonctionnels. C'est ce que montre la notion d'équilibre mixte. Par exemple, comme l'explique MAYNARD-SMITH (1984), on peut supposer qu'une population quelconque soit composée d'un certain nombre d'«éperviers» (dont la stratégie pour obtenir une ressource quelconque est de lutter à mort) et d'un certain nombre de «colombes» (qui refusent de se battre dès lors que surgit la moindre menace). Or, le caractère fonctionnel de l'un ou l'autre de ces caractères dépendra de leur proportion au sein de cette population, en vertu des équilibres mixtes : en effet, si la population est majoritairement composée d'éperviers, la stratégie adoptée par les colombes s'avère la plus fonctionnelle, puisqu'en se retirant, elles vont laisser les éperviers s'entre-tuer. Mais si les colombes sont majoritaires, alors la stratégie des éperviers s'avère nécessairement la plus fonctionnelle. La morale de la fable, ici, est simple : fonctionnalité biologique et valeur morale s'avèrent grandement hétérogènes puisqu'un comportement susceptible d'être moralement valorisé (le «pacifisme» des colombes) peut s'avérer dans certaines circonstances, hautement dysfonctionnel en termes de survie et de reproduction.

7 ➤ Conclusion

Le champ d'extension du fonctionnalisme moral naturaliste, en dernière analyse, s'avère soit trop important, soit trop restreint. Trop important lorsqu'il s'efforce de dériver la fonction morale de l'espèce humaine d'une fonction naturelle unique (comme la coopération) sans tenir compte des modalités précises d'accomplissement de cette fonction. Trop restreint lorsqu'il réduit la valeur morale d'un comportement à sa fonctionnalité en termes de survie et de reproduction (si tel était le cas, mensonges et trahisons en tous genres, hautement fonctionnels, seraient également valorisables d'un point de vue moral, ce qui bien évidemment ne va nullement de soi). Vouloir enraciner les normes morales dans l'évolution biologique, c'est méconnaître les véritables bases

évolutionnaires de notre sens moral : contrairement à ce qu'affirment certains défenseurs du réalisme moral naturaliste, ce dernier n'a pas évolué de manière à détecter un certain nombre de vérités morales, mais bien plutôt de manière à résoudre des problèmes adaptatifs qui se sont posés au cours de l'existence de nos ancêtres, problèmes qui ne sont plus nécessairement d'actualité. De sorte que le fonctionnalisme moral, loin de permettre l'édification d'un soubassement universel de la moralité, risque bien au contraire d'induire en erreur ceux qui tenteraient de mener à bien un tel projet.

Les ambitions fondationnalistes du naturalisme moral doivent par conséquent être revues à la baisse : le critère de fonctionnalité, à lui seul, ne saurait permettre de fonder une quelconque éthique normative, et ce même si les théories naturalistes peuvent par ailleurs s'avérer utiles en mettant en évidence les mécanismes psychologiques ou biologiques qui sous-tendent nos capacités morales. En effet, si le fait d'accomplir une fonction était suffisant pour assurer la validité morale d'une norme, alors la normativité morale serait toute entière dépendante de la normativité biologique. Et l'on voit mal, dans de telles conditions, comment il deviendrait possible pour la première de formuler le moindre discours critique à l'égard de la seconde. Or, il ne peut y avoir de morale que si l'on est précisément en mesure de condamner un ensemble de normes ou de tendances comportementales, si fonctionnelles soient-elles d'un point de vue biologique ou social.

À l'heure où les biotechnologies sont en passe de bouleverser notre rapport éthique au monde, il importe de mettre l'accent sur le fait que le domaine moral excède de beaucoup la nature strictement biologique de l'homme. Les questions morales contemporaines (avortement, euthanasie, clonage, etc.) ne peuvent se satisfaire d'une comparaison avec un quelconque paradigme naturaliste ancré dans l'évolution biologique et les fonctions qui s'y rattachent. Bien au contraire, elles doivent impérativement tenir compte de l'évolution culturelle et de l'évolution technologique, pièces maîtresses d'un paysage moral qu'il nous reste encore à élaborer.

☉Références bibliographiques☉

A

ARNHART Larry (1998), *Darwinian Natural Rights: The Biological Ethics of Human Nature*, State University of New York Press.

AXELROD Robert (1984), *The Evolution of Cooperation*, Basic Books.

B

BAUDOIN Cyril & BROUSSEAU Olivier (2011, à paraître), *Les Créationnistes*, Belin.

C

CLAVIEN Christine (2009), *Je t'aide, moi non plus. Biologique, comportemental ou psychologique, l'altruisme dans tous ses états*, Vuibert.

CLAVIEN Christine & EL-BEZ Catherine (dir.) (2007), *Morale et évolution biologique*, Presses polytechniques et universitaires romandes.

CUMMINS Robert (1975), "Functional analysis", *Journal of Philosophy*, 72: 741-764.

CASEBEER William (2005), *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition*, MIT Press.

D

DARWIN Charles (1992), *L'Origine des espèces* (1^{re} éd. 1859), Paris, GF.

DARWIN Charles (2000), *La Filiation de l'homme et la sélection liée au sexe* (1^{re} éd. 1871), Syllepse.

F

FAUCHER Luc (2007), «Les émotions morales à la lumière de la psychologie évolutionniste: le dégoût et l'évitement de l'inceste», in C. Clavien & C. El-Bez (dir.), *Morale et évolution biologique*, Presses polytechniques et universitaires romandes, 2007.

G

GOULD Stephen Jay (2001), *L'Éventail du vivant*, Seuil.

J

JOYCE Richard (2006), *The Evolution of Morality*, MIT Press.

M

MACKIE J.L. (1977), *Ethics: Inventing Right And Wrong*, Penguin.

MILLIKAN Ruth (1984), *Language, Thought, and other Biological Categories*, MIT Press.

MAYNARD-SMITH John (1984), "Game theory and the evolution of behaviour", *Behavioral and Brain Sciences*, 7: 95-125.

MOORE G.E. (1903), *Principia Ethica*, Cambridge University Press.

O

OGIEN Ruwen (dir.) (1999), *Le Réalisme moral*, PUF.

R

RAILTON Peter (1986), "Moral Realism", *Philosophical Review*, 95: 163-207.

S

SAYRE-McCORD G. (ed.) (1988), *Essays on Moral Realism*, Cornell University Press.

SHEA Nicolas (2006), «La contribution de Ruth Millikan à la philosophie matérialiste de l'esprit», in F. Athané, É. Machery & M. Silberstein (dir.), *Matière première. Revue d'épistémologie et d'études matérialistes*, n° 1/2006, Syllepse.

SKYRMS Brian (1996), *The Evolution of the Social Contract*, Cambridge University Press.

T

TRIVERS R.L. (1971), "The evolution of reciprocal altruism", *Quarterly Review of Biology*, 46: 35-57.

W

WILSON E.O. (1975), *Sociobiology, The New Synthesis*, Harvard University Press. Trad. fr., *La Sociobiologie*, Éditions du Rocher, 1989.

WRIGHT Larry (1976), *Teleological Explanations: An Etiological Analysis of Goals and Functions*, University of California Press.

🕒 Notice biographique 🕒

Jérôme Ravat est philosophe. Allocataire-moniteur à l'université Paris 4-La Sorbonne. Il a codirigé, avec Alberto Masala, un ouvrage sur le rapport des sciences et de la morale, à paraître fin 2010 aux Éditions Matériologiques.

Avertissement. Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur et des auteurs. Les textes et illustrations figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage personnel, scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document selon ce modèle :

Jérôme Ravat, «Fonction naturelle et fonction morale: similitude ou imposture?», in Gérard Lambert et Marc Silberstein (dir.), Matière première. Revue d'épistémologie [en ligne]. Nouvelle série, N° 1/2010 : Épistémologie de la médecine et de la santé, Éditions Matériologiques. Mis en ligne le 7 octobre 2010.

URL : www.materiologiques.com

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.



Nicolas Baumard **Une théorie naturaliste des phénomènes moraux est-elle possible ?**

Qu'est-ce que la morale ? Pourquoi les individus se comportent-ils moralement ? Sur quelles bases estiment-ils qu'une action est bonne ou mauvaise ? Il s'agit là de questions classiques en sciences humaines¹. Dans ce qui suit, je propose une approche naturaliste et mutualiste de ces questions.

L'approche naturaliste de la morale est l'héritière de la tradition du sens moral initiée par les philosophes des Lumières britanniques (Shaftesbury, Hutcheson, Butler, Smith). Par la suite, elle s'est développée avec la théorie de l'évolution puis avec la psychologie expérimentale et l'ethnologie empirique. On parle d'explication naturaliste au sens où ce type d'explication considère qu'il existe une continuité entre les sciences naturelles (biologiques et psychologiques) et les sciences humaines².

1. Cf. BOUDON (1995), DURKHEIM (1893), ELSTER (2009), WEBER (1995 [1921]).

2. Sur l'approche naturaliste appliquée à la morale, cf. en particulier ELSTER (2007), LUKES (2008), PHARO (2004b), SPERBER (1993).

L'approche naturaliste présentée ici se démarque très nettement des approches *réductionnistes* développées par la sociobiologie (WILSON 1975). De telles approches cherchent à *réduire* le niveau social et le niveau psychologique sur le niveau biologique des gènes. Une approche naturaliste cherche au contraire à *articuler* les différents niveaux entre eux (évolution, jugements, interactions sociales) (SPERBER 1997a).

En dépit de la vitalité des théories évolutionnaires et de l'importance des données empiriques tant en psychologie qu'en anthropologie, les approches naturalistes n'ont jamais réussi à l'emporter sur leurs concurrentes non naturalistes³. Dans cet article, je suggérerai que la raison de cet échec tient pour une part au peu d'intérêt des théoriciens naturalistes pour les théories normatives et descriptives développées par les philosophes moraux. Les partisans de l'approche naturaliste se sont en effet concentrés essentiellement sur le caractère naturel de notre disposition morale (autonomie, spécificité, innéité, universalité) ainsi que sur les mécanismes évolutionnaires susceptibles de l'avoir sélectionnée (altruisme réciproque, développement de la sympathie, compétition entre groupes, etc.), sans s'attacher à la logique des jugements moraux. Or c'est ce que font au contraire les philosophes moraux. Comme le note Rawls, «du moins à son stade initial, une théorie de la justice est bien une théorie. C'est une "théorie des sentiments moraux" (pour évoquer une expression du XVIII^e siècle) qui expose les principes gouvernant notre capacité morale ou, plus particulièrement, notre sens de la justice» (RAWLS 1971, p. 75).

L'approche mutualiste reprend la tradition du contrat développée par Hobbes, Rousseau, Kant et plus récemment par RAWLS (1971), GAUTHIER (1986) et SCANLON (1998)⁴. Elle considère que les relations sociales se caractérisent à la fois par un conflit d'intérêt et par une identité d'intérêts (RAWLS 1971, p. 30). Se comporter moralement, c'est viser le respect *mutuel* des intérêts de chacun (on parlera aussi d'équilibre des intérêts). Si chacune de ces théories n'est pas nouvelle, elles n'ont toutefois jamais été défendues conjointement⁵. Dans cet article, je chercherai à montrer

3. Cf. notamment les synthèses de RIDLEY (1996), WILSON (1993).

4. Il faut cependant distinguer deux traditions, la tradition *contractarienne* (Hobbes, Gauthier) qui cherche à montrer qu'il est mutuellement avantageux de suivre une morale du contrat, et la tradition *contractualiste* (Rousseau, Kant, Rawls, Scanlon) qui montre que nos intuitions morales fonctionnent comme si nous avions négocié un contrat. La première s'intéresse aux intérêts individuels, la seconde aux intuitions morales. Elles partagent cependant l'idée que se comporter moralement, c'est équilibrer les intérêts individuels.

5. Une telle approche n'est pas totalement inédite. À la fin du XIX^e siècle, le géographe et théoricien anarchiste Piotr Kropotkine en donne une esquisse dans *L'entraide: un facteur de l'évolution* (KROPOTKINE 2001 [1902]). La théorie de

que leur combinaison offre une théorie cohérente et empiriquement valide de la morale. Je présenterai d'abord les raisons qui militent en faveur d'une approche naturaliste. Dans une seconde partie, je décrirai les deux principales théories naturalistes et je montrerai leurs limites. Cela m'amènera à présenter brièvement l'alternative que constitue la théorie mutualiste, et de ses mérites. Je montrerai ensuite comment l'approche naturaliste et l'approche mutualiste peuvent se combiner en une théorie naturaliste et mutualiste de la morale. En conclusion, j'insisterai sur l'intérêt pour les sciences humaines de chercher à s'intégrer aux sciences naturelles.

La défense d'une théorie de la morale nécessiterait évidemment des développements plus conséquents (BAUMARD 2008). L'ambition de ce texte est plus modeste. Il vise à montrer l'intérêt et la plausibilité d'une approche naturaliste et mutualiste de la morale. Il cherche également à dégager les prédictions spécifiques à l'approche mutualiste et à montrer à quelles conditions cette théorie peut être réfutée. D'autres travaux, à la fois théoriques et empiriques, sont donc nécessaires. C'est ce type de recherches qu'entend susciter une telle présentation.

1 ➔ Pourquoi adopter une approche naturaliste ?

1.1 ➔ La spécificité de la morale

Il y a quelque chose de spécifique aux motivations morales. Cette spécificité constitue l'une des raisons qui militent en faveur d'une approche naturaliste (BAUMARD 2007, BAUMARD & SPERBER 2007). Considérons l'exemple classique des impôts. Je peux les payer pour des raisons strictement égoïstes, parce que je crains un redressement fiscal. Mais je

Kropotkine n'est cependant qu'une ébauche : il ne s'attarde ni sur les mécanismes évolutionnaires ni sur la structure du sens moral. [Sur Kropotkine, cf. la revue *Réfractations*, n° 23, «L'entraide, un facteur de révolutions», automne 2009, NdÉ.] Rawls, en revanche, est plus précis. Dans sa *Théorie de la justice*, il dessine les contours d'une approche naturaliste de la tradition du contrat social. Au niveau psychologique, en effet, il appelle de ses vœux la transformation de la théorie contractualiste en une théorie plus cognitive s'inspirant du travail de Chomsky sur le langage. Au niveau évolutionnaire, il suggère d'expliquer nos intuitions contractualistes à l'aide d'une théorie évolutionnaire du type de celle proposée par Trivers la même année (cf. également GAUTHIER 1986, GIBBARD 1990).

peux aussi les payer pour des raisons morales, parce que j'estime avoir le devoir de contribuer au financement des biens publics.

Ces deux raisons sont clairement distinctes. Je peux payer mes impôts pour des raisons morales en sachant que je n'ai pourtant aucune raison de craindre un contrôle fiscal. Au contraire, je peux estimer que l'impôt s'apparente au vol et, dans le même temps, juger plus prudent de payer mes impôts. Il y a là deux motivations différentes, et la première n'est pas réductible à la seconde. C'est ce qu'avait noté déjà Hutcheson. Imaginons, écrivait-il, que nous recevions le même bienfait de deux hommes différents. Le premier a agi pour des raisons morales, le second par contrainte ou pour servir ses propres intérêts. «Tous deux nous sont également avantageux, et pourtant ils nous inspirent des sentiments radicalement différents» (HUTCHESON 2004 [1725], p. 90).

De la même manière, on ne peut réduire la morale à la sympathie ou à l'amour des autres. Par exemple, le directeur d'une entreprise peut avoir de la sympathie pour un ami au chômage, et juger qu'il n'est pas moral de licencier l'un de ses employés pour recruter son ami. À l'inverse, ce même directeur peut n'avoir aucune sympathie pour l'un de ses concurrents et, dans le même temps, estimer que tous les coups ne sont pas permis pour nuire à ce concurrent.

La sympathie ne crée donc pas nécessairement un devoir. Smith examine le cas inverse. Il note que nous avons des devoirs envers quelqu'un, alors même que nous n'avons aucune sympathie pour lui.

Le soldat qui donne sa vie pour sauver celle de son officier ne serait peut-être que peu affecté par la mort de cet officier si elle se produisait sans la moindre faute de sa part, et un tout petit incident dont il pourrait par ailleurs souffrir l'affligerait sans doute bien plus vivement. (SMITH 1999 [1759], p. 266.)

Il y a donc quelque chose de spécifique à la motivation morale, qui ne se réduit ni à l'égoïsme, ni à l'altruisme⁶. C'est également le cas des jugements moraux. Pour nous en convaincre, considérons le fameux dilemme du trolley (FOOT 1967, THOMSON 1985). Dans ce dilemme, un

6. On pourrait encore évoquer la spécificité des émotions morales. Ainsi, la honte et la culpabilité ne semblent pas fonctionner de manière identiques (elles ne sont pas activées par les mêmes situations, elles ne disposent pas aux mêmes comportements, elles ne mobilisent pas les mêmes aires cérébrales – KETELAAR & AU 2003, TAKAHASHI *et al.* 2004, TANGNEY & DEARING 2002).

trolley dont le conducteur s'est évanoui fonce à vive allure vers cinq piétons se trouvant sur la voie. La seule façon de sauver les cinq piétons serait de détourner le trolley vers une autre voie. Malheureusement, un autre piéton se trouve sur cette voie. Peut-on détourner le trolley vers la voie secondaire? La plupart des personnes interrogées considèrent que oui (cf. HAUSER, CUSHMAN, YOUNG & JIN 2007, MIKHAIL, SORRENTINO & SPELKE 1998). Ce résultat est très stable et ne semble pas corrélé à l'âge, au niveau d'études, à l'appartenance religieuse (HAUSER, CUSHMAN, YOUNG & JIN 2007).

Si l'histoire s'arrêtait là, il serait aisé d'expliquer le jugement moral en termes de sympathie – cinq personnes touchent plus notre sympathie qu'une seule – ou de raisonnement – il est plus rationnel de sauver cinq vies qu'une seule. Il existe cependant une autre version du dilemme. Dans celle-ci, cinq personnes doivent toujours être sauvées. Mais cette fois, la seule façon de les sauver consiste à bloquer le trolley avec un objet lourd. Jean se trouve sur un pont piétonnier surplombant la voie, aux côtés d'un piéton portant un gros sac à dos. Jean peut-il pousser l'homme au sac à dos? Dans ce cas, la plupart des personnes interrogées estiment qu'on ne peut pas sacrifier la vie du piéton sur le pont pour sauver la vie des cinq piétons sur la voie. Le jugement moral ne semble donc pouvoir se réduire ni à la sympathie pour la peine d'autrui, ni à la maximisation rationnelle du nombre de vies sauvées. Autrement dit, les jugements moraux semblent avoir leur propre logique, une logique qui leur est spécifique. Une manière d'expliquer cette spécificité consiste à postuler l'existence d'un mécanisme psychologique spécifique. Une telle hypothèse relève d'une approche naturaliste dans la mesure où le phénomène moral est expliqué à l'aide d'un mécanisme compatible avec les sciences naturelles.

1.2 ↻ Le sens moral

Comme je l'ai indiqué en introduction, l'approche naturaliste n'est pas nouvelle. Pour les philosophes du sens moral, nous sommes équipés d'un sens moral, organe mental chargé de la morale, tout comme le cœur, organe physique, est chargé de la circulation sanguine.

Au premier abord, une telle approche peut sembler paradoxale. Examinons rapidement quatre types d'objections. La première objection consiste à observer que le sens moral n'est pas compatible avec le comportement immoral. Si nous sommes naturellement moraux, pour-

quoi commettons-nous des fautes? Il n'y a là rien de contradictoire. Comme le dit Butler, la morale est une passion parmi d'autres (BUTLER 2008 [1726]). Je me sens le devoir de voter mais, dans le même temps, je préférerais profiter du jour du vote pour partir au bord de la mer. Deux motivations sont ici en conflit, et la motivation morale ne gagne pas toujours. Je reste parfois chez moi, alors que j'estime que je devrais aller voter.

Seconde objection: la morale ne peut être le produit automatique d'un organe mental. Elle s'appuie en effet sur l'autonomie, la réflexion, les raisonnements de chaque individu, et a partie liée à la philosophie, au droit, à la religion. Ces observations ne sont pas antinomiques avec l'existence d'un sens moral. D'une part, l'existence d'un organe moral est parfaitement compatible avec l'autonomie morale. Si chaque individu est pourvu d'un sens moral, c'est chaque individu, avec son histoire, son éducation, les informations dont il dispose, la situation dans laquelle il est, qui porte un jugement, à l'aide des procédures de justice et de morale implémentées dans son sens moral.

D'autre part, l'existence de processus automatique ne contredit pas celle de processus plus théoriques. Les humains ont ainsi des intuitions mathématiques sur les quantités et l'espace (DEHAENE 1997). Dans le même temps, de nombreuses cultures ont développé des systèmes de comptage et des théories mathématiques élaborées. Ces deux types de processus, intuitifs et théoriques, coexistent sans pour autant se mélanger (EVANS 2008).

Il se peut ainsi qu'à côté de nos réflexions éthiques, nous utilisons dans la vie quotidienne un sens moral plus automatique (pour une revue de cette question, cf. HAIDT 2001). Nos jugements moraux sont *rapides* – quelques dixièmes de secondes –, *automatiques* – nous ne pouvons nous empêcher de considérer une situation d'un point de vue moral – et *intuitifs* – nous ignorons les processus cognitifs qui nous ont conduit à un jugement particulier, nous ne disposons que du résultat (l'intuition) de ces processus cognitifs⁷. En d'autres termes, le sens moral fonctionne de manière *autonome*.

7. Intuitif s'oppose ici à réflexif (EVANS 2008, SPERBER 1997b). Ainsi, une démonstration mathématique est réflexive parce que nous en connaissons à la fois les étapes et le résultat.

Nous «sentons» donc le bien et le mal, plus que nous y réfléchissons. Voilà l'analogie entre sens moral et capacités sensorielles. Tout comme l'ouïe ou la vue, notre disposition morale fonctionne de façon rapide, automatique et inconsciente. Ainsi, nous ne pouvons pas davantage nous empêcher de considérer le monde d'un point de vue moral que nous ne pouvons éviter de le percevoir en couleurs. Et, tout comme nous sentons qu'un son est aigu sans rien connaître de l'acoustique, nous «sentons» qu'une action est condamnable avant même d'y avoir réfléchi. Les études sur le dilemme du trolley en offrent une illustration. La majeure partie des participants sont ainsi incapables d'expliquer pourquoi ils acceptent de détourner le trolley vers le piéton se trouvant seul sur les rails, mais refusent de pousser l'homme sur le pont (HAUSER, CUSHMAN, YOUNG & JIN 2007)⁸.

Troisième objection : les jugements moraux sont très divers et varient d'une société à l'autre. L'infanticide est sans doute l'un des exemples les plus extrêmes de divergence morale. Il arrivait ainsi aux Inuits de tuer leurs enfants en bas âge. Rien ne saurait être plus contraire à notre morale. Comment donc un sens moral universel peut-il produire des jugements si différents ? Comme l'avait noté Adam Smith, cette objection n'a rien de définitif :

Il est fréquemment impossible de nourrir à la fois son enfant et soi-même. Nous ne pouvons dans ce cas nous étonner qu'il l'abandonne. Celui qui, fuyant un ennemi auquel il est impossible de résister, devrait laisser son enfant qui retarde sa fuite, serait certainement excusable puisqu'en essayant de le sauver, il ne pourrait guère espérer que la consolation de mourir avec lui. (SMITH 1999 [1759], p. 291).

Les différences peuvent non seulement provenir du contexte mais également des croyances factuelles sur lesquelles se fondent nos jugements. Considérons le débat sur la redistribution. Américains et Européens s'accordent sur le devoir d'assistance envers les plus démunis mais s'opposent sur l'importance de l'aide à accorder. Or une majorité d'Américains estiment que les pauvres sont responsables de leur pauvreté, tandis que la plupart des Européens pensent qu'ils ne peuvent

8. Une telle conclusion n'implique pas qu'il n'existe pas de réflexions morales. Il existe quantité de réflexions morales, qui peuvent jouer un rôle déterminant. Il n'en reste pas moins qu'une grande partie des jugements moraux ne s'appuie pas effectivement sur des processus réflexifs.

s'extraire de la pauvreté par eux-mêmes (notons que les États-Unis et l'Europe ont des niveaux comparables de mobilité sociale, cf. ALESINA & GLAESER 2004). Ces croyances factuelles à propos de la société expliquent sans doute en partie la plus grande tolérance des Américains à l'inégalité (pour d'autres exemples, cf. BOUDON 1995, LUKES 2008, SPERBER 1993).

La dernière objection concerne le caractère inné du sens moral. Comment le sens moral pourrait-il être inné ? D'une part, nous ne naissons pas moraux. D'autre part, les enfants n'ont pas les mêmes jugements que les adultes. Shaftesbury répond à la première objection de la manière suivante : « Si vous n'aimez pas le mot "inné", changez de terme. Si vous le voulez, dites "instinct" et appelez instinct tout ce que la nature enseigne » (SHAFTESBURY 1714). Inné signifie ici un mécanisme psychologique dont le développement est plus ou moins contraint par la génétique. Il ne signifie pas en revanche que ce mécanisme est présent dès la naissance.

Il est certes vrai que les enfants n'ont pas les mêmes jugements que les adultes. Cela n'a cependant rien d'un contre-argument. Il se peut que les enfants aient les mêmes principes que les adultes, mais les appliquent différemment parce qu'ils évaluent la situation d'une façon différente des adultes. Par exemple, les enfants ne saisissent pas toujours la subtilité des rapports sociaux et peuvent blesser un camarade sans s'en rendre compte. Ainsi, des actions qualifiées d'immorales par les adultes, qui en perçoivent le caractère blessant, sont jugées acceptables par les enfants les plus jeunes.

L'évolution des jugements moraux au cours du développement individuel ne constitue donc pas un contre-argument à l'idée que les humains sont naturellement moraux. Au contraire, de nombreuses études de psychologie morale témoignent du caractère naturel de la morale. Premièrement, les jugements moraux apparaissent très tôt, dès avant 2 ans. Dès avant 2 ans, les enfants jugent ainsi une action en fonction du tort qu'elle cause (DUNN 1988). Au même âge, les enfants s'excusent en protestant de leur innocence (« Je ne l'ai pas fait exprès », « Je ne savais pas ») ou par des circonstances atténuantes (« C'est lui qui a commencé ») entend-on dans la cour de l'école maternelle).

Deuxièmement, la logique des jugements moraux ne semble pas changer au cours du temps. Les enfants de maternelle évaluent une faute morale en fonction du tort causé à autrui (TURIEL 1983) et font la

différence entre un comportement immoral et un comportement imprudent (TISAK & TURIEL 1984). Ils font la différence entre une souffrance méritée et une souffrance indue (LESLIE, MALLON & DICORCIA 2006). Ils distribuent le fruit d'une action commune en fonction de la contribution de chacun (BAUMARD, CHEVALLIER & MASCARO, soumis). Les justifications s'améliorent au cours du temps, certes, mais les principes ne changent pas⁹.

Troisièmement, les enfants ne semblent pas se baser sur ce que leur enseignent les adultes. Comme le notait Piaget, l'enseignement des adultes est souvent trop général pour être utile. «Les parents moyens sont comme des gouvernements inintelligents qui se bornent à accumuler des lois, au mépris même des contradictions et de la confusion d'esprit croissante qui résulte de cette accumulation» (PIAGET 1932, p. 152; pour une étude empirique, cf. Ross 1996). De fait, lorsqu'ils doivent justifier leurs jugements moraux, les enfants se contentent rarement d'évoquer des règles qui auraient pu leur être inculquées (DAMON 1988). Ils invoquent plutôt les paramètres moraux qui leur semblent pertinents (responsabilité, proportionnalité, etc.). Si les adultes jouent un rôle dans le développement du jugement moral, ce n'est donc pas, semble-t-il, au niveau des *principes moraux*, mais au niveau de leur *application concrète*

9. Les premières études sur le développement de la morale semblaient pourtant montrer que nos jugements moraux ne s'appuient pas sur une capacité innée (PIAGET 1932). En réalité, ces études n'ont pas tant porté sur les intuitions morales que sur les réflexions morales. Ainsi, au début du *Jugement moral chez l'enfant*, Piaget se donne pour but de comprendre «la conscience de la règle, c'est-à-dire la manière dont les enfants de différents âges se représentent le caractère obligatoire, sacré ou décisive, l'hétéronomie ou l'autonomie propre aux règles du jeu [de billes]» (*ibid.*, p. 2). Ce qui intéressait Piaget, c'est ce que pensent les enfants des règles morales : d'où viennent-elles ? Sont-elles intemporelles ? Reposent-elles sur l'accord mutuel ? Il oppose ainsi les jeunes enfants qui perçoivent les règles morales comme intangibles et dictées par les adultes ou la tradition, aux enfants plus âgés jugeant que ces règles dérivent de propositions individuelles acceptées par tous. Dans l'esprit de Piaget, il s'agit clairement de questions théoriques. Il qualifie par exemple les questions qu'il pose aux enfants sur l'origine des règles morales de questions portant sur leur «philosophie de l'histoire». La même remarque peut-être faite du travail de Kohlberg sur l'évolution des justifications (et non des jugements) chez les enfants (KOHLBERG 1969).

(La personne souffre-t-elle? Ce partage est-il proportionnel?, etc.). Les enfants ne paraissent pas tenir leurs principes moraux des adultes¹⁰.

Pour conclure, l'existence d'un sens moral n'est pas incompatible avec la richesse des phénomènes moraux : l'égoïsme, la diversité culturelle et interpersonnelle des jugements, leur autonomie du jugement moral, les réflexions éthiques, l'évolution des jugements avec l'âge ne constituent pas des objections à l'approche naturaliste.

1.3 ➔ Le sens moral, une adaptation

À l'époque des philosophes moraux, l'idée de sens moral rencontrait cependant un problème fondamental. Comment expliquer que *tous* les hommes soient équipés d'un *même* sens moral? Ce problème n'avait rien de spécifique au sens moral. Il se posait également pour les autres capacités psychologiques, ainsi que pour les capacités physiques : comment expliquer l'œil, comment expliquer la main? Faut-il invoquer la volonté d'un ingénieur suprême? Ce problème se trouvera résolu au XIX^e siècle avec la théorie de la sélection naturelle proposée par Darwin en 1859 dans *On the Origin of Species by Means of Natural Selection (L'Origine des espèces)*. Par la suite, en 1871, Darwin consacre d'ailleurs de nombreuses pages à l'émergence du sens moral dans *The Descent of Man, and Selection in Relation to Sex (La Descendance de l'homme)*. Selon la théorie de la sélection naturelle, il est possible qu'un sens moral ait été sélectionné si, dans l'environnement ancestral, il était avantageux (pour les individus ou pour les groupes) de se comporter moralement.

La théorie de l'évolution offre donc un cadre scientifique pour expliquer l'existence d'un sens moral. Elle éclaire également les caractéristiques que nous avons rencontrées plus haut : spécificité, automaticité,

10. Une telle analyse n'implique pas que toutes nos intuitions soient le produit d'un mécanisme inné. Notre capacité à lire est également automatique, rapide et inconsciente. Pour autant, elle n'est pas le produit de la sélection naturelle. Notons cependant que la lecture présente une grande uniformité à travers les cultures et les systèmes d'écriture (en terme d'apprentissage, de fonctionnement, de localisation cérébrale). Cette uniformité s'explique par le fait que la lecture détourne à son profit des capacités plus basiques (reconnaissance des formes, mémoire visuelle, boucle phonologique) qui elles sont innées (DEHAENE 2008). Nature et culture ne sont donc pas antinomiques. Au contraire, leur combinaison est souvent nécessaire à la compréhension d'un phénomène (SPERBER & HIRSCHFELD 2004).

innéité. Ces caractéristiques contribuent à la fonctionnalité du sens moral. En effet, s'il est avantageux aux individus d'être moraux, il est dans leur intérêt d'être préparés de manière innée à se comporter moralement (tout comme ils sont préparés à marcher ou à parler). De la même manière, il est avantageux de disposer d'un système spécifique et autonome. Comme l'observe HUTCHESON (2004 [1725], p. 9), si nous devons utiliser des capacités générales de raisonnement chaque fois que nous sommes dans une situation morale, nous ne serions pas assez rapides et nous ferions trop d'erreurs. Plus précisément, l'avantage de la spécificité et de l'autonomie dépend du type de problème à traiter. Si celui-ci présente des caractéristiques particulières et récurrentes et s'il demande une réponse elle aussi particulière, il peut être avantageux, pour l'organisme, de développer un dispositif spécialisé pour ce type de problèmes. Un tel dispositif composé de routines et de raccourcis propres au problème à traiter sera plus efficace qu'un mécanisme plus général qui doit, à chaque fois, retrouver la solution du problème¹¹.

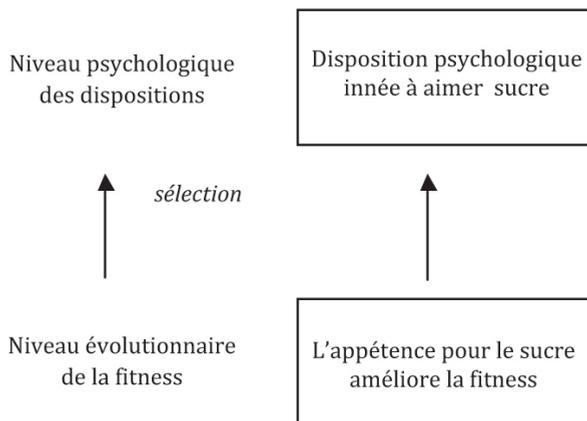
Encore une fois, de telles caractéristiques ne sont pas propres au sens moral (pour une revue de la littérature sur le sujet, cf. PINKER 1997). Évoquons rapidement le cas du dégoût. L'autonomie du dégoût se révèle de manière spectaculaire par notre incapacité à manger des aliments que nous savons être sans danger. Si un verre a précédemment contenu un cafard, nous refuserons de boire ce qu'il contient même s'il a été lavé et stérilisé devant nous (ROZIN, HAIDT & McCAULEY 1993). Le sens du dégoût fonctionne de manière spécifique. Par exemple, il n'est pas sensible aux effets de proportionnalité. Le simple contact d'un objet avec

11. L'avantage de la spécialisation n'est pas propre à l'évolution. De fait, en biologie comme en ingénierie, les systèmes les plus efficaces ne sont pas les systèmes les plus généraux, mais au contraire les systèmes composés de modules spécifiques et autonomes (CALLEBAUT & RASSKIN-GUTMAN 2005, SCHOLL & LESLIE 1999). Le degré de spécificité du dispositif est une affaire de calibrage. Celui-ci dépend entièrement des caractéristiques du problème, des ressources biologiques disponibles, de la pression de sélection. Les humains sont ainsi équipés d'un dispositif spécialisé dans la seule détection des serpents (OHMAN, FLYKT & ESTEVES 2001). L'existence de ce dispositif s'explique à la fois par le danger important représenté par les serpents dans l'environnement ancestral, et par les caractéristiques visuelles des serpents (mouvement sinusoïdal) qui rendent aisée la sélection d'un dispositif spécialisé. Au final, le calibrage est effectué par la sélection naturelle, problème par problème.

un autre objet dégoûtant peut suffire à rendre le premier également dégoûtant. Ce mécanisme s'explique par la *fonction* du sens du dégoût. Celui-ci a en effet été sélectionné pour nous prévenir des substances contaminantes contenant des pathogènes comme les bactéries, les virus, les micro-organismes dont le seul contact peut être dangereux.

1.4 ⇨ Niveau évolutionnaire et niveau psychologique

Si la morale a évolué pour l'avantage des individus ou des groupes, cela signifie-t-il qu'en agissant moralement, nous visons notre intérêt propre ou celui de notre groupe? Rien ne serait plus faux. Considérons pour le comprendre notre goût pour le sucre (cf. fig. 1, ci-dessous). Il était dans notre intérêt de manger des aliments sucrés, comme les fruits. Par ailleurs, nous apprécions le goût sucré pour lui-même. Tout se passe donc *comme si* nous mangions des aliments sucrés pour leurs qualités nutritives. Ce n'est pourtant pas le cas. Il n'est pas besoin de connaître le rôle des glucides pour avoir envie de manger sucré. Ce rapport entre les qualités nutritives des glucides et notre appétence pour les aliments qui en contiennent le plus n'a évidemment rien de fortuit. Il s'agit d'une adaptation à l'environnement ancestral (où la production industrielle de sucre était absente). Dans le passé de notre espèce (et de nombreuses autres qui apprécient le goût du sucre), les individus présentant une disposition innée à aimer le sucré choisissaient des aliments plus riches en énergie. Ils vivaient donc mieux, plus longtemps et avaient plus de descendants (en termes évolutionnaires, leur fitness était plus élevée). À leur tour, ces descendants possédaient ce goût pour le sucré.



On peut donc analyser un même phénomène (notre goût pour le sucre) à deux niveaux, au niveau évolutionnaire (la fitness) et au niveau psychologique (les jugements). Au niveau évolutionnaire, nous avons une disposition à manger du sucre parce que le sucre apportait de l'énergie à nos ancêtres – parce que manger du sucre améliorait leur fitness. Au niveau psychologique, nous mangeons du sucre parce que nous en aimons le goût – parce que nous jugeons que le sucre est agréable.

Appliquons à présent cette distinction à la morale. Si la théorie évolutionnaire que nous avons esquissé se vérifie, il y avait, *au niveau évolutionnaire*, un intérêt à se comporter moralement. Les individus ou les groupes contenant des membres pourvus d'une disposition morale étaient avantagés. Ils réussissaient mieux que les autres individus ou que les autres groupes. *Au niveau psychologique*, les humains n'agissent pas, cependant, avec pour objectif leur intérêt propre ou celui de leur groupe. Ils ne font que chercher à agir le plus moralement possible.

La perspective évolutionnaire permet d'expliquer l'existence d'un dispositif spécifiquement dédié à la morale. Il reste cependant à préciser les mécanismes évolutionnaires qui ont pu conduire à l'émergence d'un sens de la morale. En d'autres termes, quelle est la fonction du sens moral? Dans *La Descendance de l'homme*, Darwin a proposé deux théories naturalistes: la théorie continuiste et la théorie altruiste. Ces deux théories sont encore aujourd'hui les théories dominantes. Nous allons voir cependant qu'elles ne semblent pas pouvoir rendre compte de l'ensemble des phénomènes moraux. C'est la raison pour laquelle nous serons amené à explorer une théorie naturaliste alternative, inspirée de l'approche mutualiste.

2 ⇨ Quel est l'intérêt de l'approche mutualiste ?

2.1 ⇨ La théorie continuiste

Présentons la théorie continuiste avec les mots de Darwin:

La proposition suivante me paraît avoir un haut degré de probabilité: un animal quelconque, doué d'instincts sociaux prononcés en comprenant, bien entendu, au nombre de ces instincts, l'affection des parents pour leurs enfants et celle des enfants pour leurs parents, acquerrait inévitablement un sens moral ou une conscience, aussitôt que ses facultés intellectuelles se seraient développées aussi complètement ou presque aussi complètement qu'elles le sont chez l'homme. En effet, les instincts sociaux poussent l'animal à trouver du plaisir dans la société

de ses semblables, à éprouver une certaine sympathie pour eux, et à leur rendre divers services. (DARWIN 1891 [1871], p. 104.)

Selon la théorie continuiste, la morale se place dans la *continuité* des instincts déjà existants. C'est de la sympathie, de l'instinct parental, du dégoût, du respect de la hiérarchie avec en plus de la raison, de la culture, du langage.

Notre «sens moral» ou notre conscience, se compose d'un sentiment essentiellement complexe, basé sur les instincts sociaux, encouragé et dirigé par l'approbation de nos semblables, réglé par la raison, par l'intérêt, et dans des temps plus récents, par de profonds sentiments religieux, renforcés par l'instruction et l'habitude. (*Ibid.*, p. 143.)

Cette théorie a été largement reprise à la fois en biologie (DE WAAL 1996, HINDE 2002), en psychologie (HAIDT & JOSEPH 2004, NICHOLS 2004) et dans les sciences sociales (WILSON 1993). Selon cette théorie, être moral, c'est posséder certaines vertus. Le type de morale prédite par la théorie continuiste correspond à ce que l'on appelle en éthique normative la *morale des vertus* (HURSTHOUSE 2007, MACINTYRE 1981). L'éthique des vertus met en effet l'accent non pas sur le contenu de devoir, mais sur les dispositions psychologiques nécessaires à son accomplissement. «Il ne s'agit pas tant, pour être une personne morale, de connaître les bons principes moraux que de posséder les bonnes dispositions – par exemple, la capacité de percevoir les besoins des individus et d'y répondre de façon imaginative» (KYMICKA 1990, p. 286). Être vertueux, c'est faire usage de la bonne disposition au bon moment (compatir quand les autres sont dans la détresse, être courageux dans le danger, etc.). Pour faire son devoir, il faut donc cultiver les vertus (compassion, courage, etc.). Ainsi, un ami dépourvu de sympathie ou une mère sans affection pour ses enfants ne peuvent remplir leurs devoirs d'ami ou de parent.

En dépit de sa simplicité, la théorie continuiste ne permet pas de rendre compte des intuitions morale. Comme je vais le montrer brièvement, les intuitions non morales n'ont pas la même logique que les intuitions morales. Par conséquent, les mécanismes psychologiques non moraux ne peuvent pas expliquer la morale. Examinons trois exemples, celui de la sympathie, de l'instinct parental et du dégoût. La sympathie paraît intuitivement liée à la morale. De fait, sympathie et morale nous conduisent toutes les deux à nous tourner vers les autres. Pourtant, comme nous l'avons vu, sympathie et morale ne visent pas la même chose. Pour prendre un autre exemple, on peut avoir de la sympathie

envers un criminel et juger qu'il doit être puni. Enfin, on peut avoir de la sympathie pour quelqu'un qui se trouve dans la détresse et avoir le sentiment que l'on n'a pas le devoir de l'aider. Autrement dit, la logique de la morale et la logique de la sympathie diffèrent l'une de l'autre.

De la même manière, l'instinct parental peut entrer en conflit avec le sens moral. C'est le cas, par exemple, lorsqu'un parent aimerait enfreindre le règlement d'un examen pour aider son enfant : son affection pour son enfant le pousse à tricher, mais son sens moral le retient de commettre une injustice. À l'inverse, un parent a des devoirs envers ses enfants, même s'il n'a pas d'affection pour eux. Là encore, la logique de l'instinct de parenté et celle de la morale ne sont pas les mêmes.

Terminons par le dégoût. Dégoût et condamnation morale vont souvent de pair. Pour autant, ils peuvent également entrer en conflit. Je peux ainsi être dégoûté par certaines pratiques sans pour autant les condamner au non du respect de la liberté de chacun (cf. par exemple HAIDT & HERSH 2001).

Pour conclure, la morale ne semble pas s'inscrire aisément dans la continuité des sentiments non moraux. Devoir aider quelqu'un, ce n'est pas la même chose qu'avoir envie de l'aider. Devoir aimer ses enfants, cela n'équivaut pas à aimer ses enfants. Condamner une action dégoûtante, ce n'est pas seulement être dégoûté par cette action. L'éthique des vertus ne suffit donc pas à caractériser la logique de nos intuitions morales. Certes, les vertus nous aident à accomplir notre devoir. Mieux vaut être sympathique si l'on veut être un bon ami. Pour autant, la sympathie peut être utilisée à mauvais escient, elle peut conduire à l'indulgence ou à la complaisance envers ses amis. Autrement dit, la sympathie n'est pas intrinsèquement morale. La théorie continuiste ne peut pas expliquer pourquoi la sympathie est parfois une vertu (compassion) et parfois un vice (complaisance). Pour conclure, les intuitions de sympathie, de dégoût, ou de parenté n'ont pas la même logique que les intuitions morales. Les mécanismes psychologiques non moraux ne suffisent donc pas à rendre compte des intuitions morales.

2.2 ➔ La théorie altruiste

Tournons-nous à présent vers l'autre théorie naturaliste, la théorie altruiste. Reprenons là encore les mots de Darwin.

Si une tribu renferme beaucoup de membres qui possèdent à un haut degré l'esprit de patriotisme, de fidélité, d'obéissance, de courage et de sympathie, qui sont toujours prêts, par conséquent, à s'entraider et à

se sacrifier au bien commun, elle doit évidemment l'emporter sur la plupart des autres tribus; or c'est là ce qui constitue la sélection naturelle. (DARWIN 1891 [1871], p. 143.)

La théorie proposée par Darwin correspond à ce que l'on appelle aujourd'hui la théorie de la sélection de groupe (BOYD, GINTIS, BOWLES & RICHERSON 2003, SOBER & WILSON 1998). Selon cette théorie, la morale a été sélectionnée pour améliorer la fitness du groupe. Les théories de la sélection de groupe partent de l'idée que les groupes sont en compétition les uns avec les autres. Dans ce contexte, les groupes qui comptent des individus disposés à se sacrifier pour le groupe surpasseront les autres groupes. À l'intérieur du groupe, les individus disposés à se sacrifier sont exploités par les autres. Cependant, dans la mesure où les groupes comptant le plus d'individus disposés à se sacrifier survivent mieux, la disposition à se sacrifier se répand parmi l'espèce humaine. Les individus altruistes laissent donc moins de descendance. Cependant, leur descendance prospère mieux dans la mesure où elle se trouve, en moyenne, dans un groupe qui surpasse les autres grâce à l'altruisme de ses membres. Autrement dit, ce que le gène de l'altruisme perd dans la compétition interindividuelle au sein du groupe se trouve compensé dans la compétition entre groupes.

Comme le notait déjà Darwin, la théorie altruiste prédit la sélection d'une morale qui vise le « bien général ou la prospérité de la communauté » (DARWIN 1891 [1871], p. 130). Il s'agit là de ce que l'on appelle en éthique normative une *morale utilitariste* (KYMICKA 1990; les économistes parlent aussi de « welfarisme », cf. par exemple SCHOKKAERT 1999). Pour l'utilitarisme, est moral ce qui maximise le bien non des individus mais de la collectivité. Jonathan Haidt oppose ainsi « l'approche contractualiste et l'approche de la ruche » (HAIDT 2007a, 2007b).

L'approche de la ruche, au contraire, considère le groupe et son territoire comme des sources fondamentales de valeur. Les abeilles, en tant qu'individus, naissent et meurent par milliers, mais la ruche leur survit, et chaque individu a un rôle à jouer pour assurer le succès de celle-ci. Les deux problèmes fondamentaux de la vie sociale sont les attaques de l'extérieur et la contestation de l'intérieur. Un seul d'entre eux peut conduire à la mort de la ruche. Tous doivent donc s'unir, faire leur devoir et être prêts à se sacrifier pour le groupe. (HAIDT 2007a.)

Si nous voulons évaluer la théorie altruiste, il convient donc de se demander si la morale vise effectivement le bien de la communauté. Nous allons donc nous poser la question suivante: les humains ont-ils

une morale utilitariste? Considérons trois exemples : la justice distributive, la justice rétributive et la souffrance induite.

2.2.1 ➔ La justice distributive

Il existe un grand nombre de données empiriques sur les jugements de justice distributive : enquêtes statistiques (MARSHALL, SWIFT, ROUTH & BURGOYNE 1999), entretiens (DUBET 2006), expériences sur la juste répartition des salaires (MICHELBAACH, SCOTT, MATLAND & BORNSTEIN 2003, MITCHELL, TETLOCK, MELLERS & ORDÓÑEZ 1993). Toutes attestent que la grande majorité des gens estiment que la distribution de la richesse ne doit pas viser la seule maximisation du bien-être global mais doit se faire en fonction de la contribution de chacun à la société. Si des gens s'indignent ainsi de ce que les salaires des PDG ou les stars de football soient aussi élevés, ce n'est pas, ou pas seulement, parce qu'ils jugent de tels salaires inefficaces d'un point de vue économique, mais parce qu'ils estiment que de telles sommes ne sont pas *méritées*.

Cette intuition est-elle incompatible avec la théorie utilitariste ? Peut-être pas. En effet, les gens pourraient juger qu'une répartition inégalitaire crée des incitations (à étudier, à travailler plus) qui profitent à l'intérêt général. Cependant, les personnes interrogées justifient moins souvent l'inégalité par l'incitation ou l'efficacité que par le mérite. Elles jugent ainsi que l'on doit donner plus à ceux qui méritent plus, peu importe l'utilité ou l'efficacité de cette répartition.

Pour tester l'importance du mérite, Mitchell et ses collègues (MITCHELL, TETLOCK, NEWMAN & LERNER 2003) ont demandé à des participants de choisir entre plusieurs types de sociétés (plus ou moins inégalitaires, plus ou moins riches), tout en faisant varier la méritocratie de chacune de ces sociétés (le rapport entre l'effort des individus et ce qu'ils obtenaient). Dans une perspective utilitariste, le mérite n'a d'importance que s'il augmente la richesse globale (par exemple en incitant les individus à travailler plus). Les réponses des participants montrent qu'ils conçoivent le mérite d'une manière indépendante de la maximisation de l'utilité globale. Les participants jugent ainsi préférable une société méritocratique plus pauvre à une société non méritocratique plus riche. De même, à richesse égale, plus la société est dite méritocratique (c'est-à-dire plus le revenu est corrélé à l'effort) et plus les participants acceptent l'inégalité (qui contribue pourtant sans doute à décroître le bien-être global de la société). L'interprétation utilitariste (rémunérer au mérite augmente la productivité) ne tient donc pas.

Ces intuitions concernant la distribution des rémunérations se retrouvent au niveau, plus général, de l'évaluation des politiques publiques. Ainsi, nous pouvons reconnaître que certaines réformes augmentent le bien-être global et néanmoins les juger inacceptables. Comme le note RAWLS (1971, p. 53) : « Nous pensons que chaque membre de la société possède une inviolabilité fondée sur la justice [...] qui a priorité sur tout, même sur le bien-être des autres. »

Jonathan Baron s'est par exemple intéressé aux jugements moraux dans la situation suivante. Imaginons deux types de patients atteints par la même maladie. On utilise un traitement qui permet de sauver 50 % des patients de chaque groupe. Un autre traitement est possible, qui permet un taux de guérison de 80 % pour le groupe A, mais réduit le taux de guérison du groupe B à 40 %. La majorité des participants refusent de changer de traitement même lorsque les médecins n'ont pas les moyens de connaître le type de chaque patient et même quand le groupe A est plus nombreux que le groupe B. Autrement dit, les participants refusaient de diminuer le bien-être de certains pour augmenter le bien-être global (BARON 1995).

2.2.2 ➔ La justice rétributive

Examinons à présent le cas de la justice rétributive. La punition vise-t-elle le bien du groupe ? En juin 2008, la Cour suprême des États-Unis a rejeté l'avis de la cour suprême de Louisiane, qui autorisait la mise à mort des personnes reconnues coupables du viol d'un enfant. La décision a été rendue à une majorité de cinq contre quatre en faveur de l'invalidation, les juges majoritaires estimant que la peine capitale ne devait être appliquée que pour les meurtriers ou les auteurs de crimes contre l'État. « La peine de mort n'est pas une punition proportionnée pour les viols d'enfant », a justifié le juge A. Kennedy. Au contraire, les deux candidats à la présidentielle américaine de 2008 ont critiqué le jugement, estimant que le viol d'enfant est un crime suffisamment odieux pour mériter la peine de mort. Dans ce débat entre la Cour suprême et les deux candidats, nulle considération n'a été accordée à l'utilité de la peine de mort, par exemple à sa capacité à dissuader les crimes futurs. Dans les deux camps, la question était bien d'évaluer dans quelle mesure la peine était juste, compte tenu du crime qui avait été commis.

L'intuition mise au jour dans cet exemple est largement confirmée par des enquêtes par questionnaires montrant elles aussi que l'utilité de la peine de mort ne semble pas être pertinente dans les débats. Plus

frappant encore, les adversaires de la peine de mort ne changeraient pas d'avis s'il leur était démontré que la peine de mort peut prévenir d'autres crimes. Inversement, les partisans de la peine de mort continueraient à soutenir cette mesure même si des études prouvaient que la peine de mort est complètement inutile en termes de dissuasion (TYLER & WEBER 1983). Pour les conservateurs, la peine de mort est juste : elle prend au meurtrier la vie qu'il a volée. Pour les libéraux, la peine de mort est injuste : elle touche de manière disproportionnée les Noirs ou les pauvres.

Les études expérimentales ont examiné les paramètres pris en compte dans les jugements de justice rétributive. Selon l'utilitarisme, la punition doit viser le bien-être de la société, par exemple en dissuadant les criminels potentiels. L'une des propositions des utilitaristes consiste par exemple à punir davantage les crimes qui sont relativement difficiles à détecter. Du point de vue du criminel, la faible probabilité de se faire prendre sera compensée par l'importance de la punition. Symétriquement, si la punition n'a aucun effet dissuasif sur certains crimes (parce que, par exemple, les criminels ne pensent pas à la prison lorsqu'ils commettent ce type de crime), il est alors inutile de laisser le criminel souffrir en prison. Cela ne ferait que diminuer encore un peu plus le bien-être global (le criminel souffrirait pour rien en prison et l'argent dépensé à l'y garder serait détourné d'usages plus intéressants). Si les jugements à propos de la punition sont utilitaristes, les individus jugeront approprié de punir durement les crimes difficiles à détecter.

Les intuitions réelles correspondent-elles à cette prédiction ? De nombreuses expériences ont examiné, dans des affaires très différentes (faute médicale, contrôle fiscal, licenciement abusif, protection du consommateur, environnement), les intuitions des participants sur le montant de la peine ou de l'amende. Considérons deux paramètres pertinents pour une approche utilitariste. Premièrement, il convient de punir davantage si l'augmentation de la punition prévient efficacement les crimes futurs. Deuxièmement, la punition doit tenir compte de la fréquence de chaque type de crime : si les crimes se font plus fréquents, on pourra estimer que cette tendance doit être contrecarrée par un renforcement de la punition. Les résultats de ces expériences indiquent que les participants ne sont pas sensibles aux paramètres utilitaristes. Par exemple, les participants refusent de ne pas punir une entreprise pharmaceutique qui aurait commis des erreurs mineures même lorsqu'ils savent que la punition entraînera l'arrêt de la production

d'un vaccin ou d'un contraceptif qui profite à l'ensemble de la société¹². Même les étudiants en droit de l'université de Chicago, université qui se distingue par la place accordée à l'approche utilitariste de la punition, refusent de tenir compte des paramètres utilitaristes. Si, pour des raisons budgétaires, les contrôles fiscaux sont moins fréquents en Utah qu'en Californie, 75 % des étudiants interrogés s'opposent néanmoins à la possibilité de punir plus sévèrement les fraudes fiscales en Utah qu'en Californie (SUNSTEIN, KAHNEMAN & SCHKADE 1998). En matière de punition, notre intuition ne vise pas le bien de la communauté.

2.2.3 ➔ Les devoirs envers les autres

On peut enfin se demander si les devoirs envers les autres visent effectivement le bien du groupe? Nous avons certes le devoir de venir en aide au blessé sur le bord de la route, mais ces devoirs ont des limites. Mais nous n'avons pas le devoir de nous dépouiller de notre argent même au profit de plus pauvre que nous. Voilà pourtant une action qui améliorerait sensiblement le bien de la société. Elle est pourtant au-delà du devoir. Elle est surrogatoire.

L'existence d'actions surrogatoires est en contradiction avec la morale utilitariste (comme le reconnaissent les philosophes utilitaristes, cf. par exemple KAGAN 1989). En effet, si la morale vise le bien-être global, alors tant que le coût de mon action génère un bénéfice supérieur pour l'ensemble de la communauté, cette action est moralement requise. Ce n'est pourtant pas ce que l'on observe. La notion de surrogatoire est très largement partagée à travers les cultures (HEYD 1982) et relativement intuitive puisque les enfants considèrent très tôt que voler n'est pas acceptable, mais que s'abstenir de donner dans la rue n'est pas moralement condamnable (KAHN 1992).

Pour conclure, ni la théorie altruiste, ni la théorie continuiste ne paraissent donc être de bonnes théories de la morale. La raison de cet échec tient sans doute pour une bonne part à la façon dont elles procèdent. Ces théories naturalistes partent des mécanismes évolutionnaires (complexification des sentiments non moraux, sélection de groupe) et en déduisent une théorie du sens moral. Dans ce qui suit, je propose une démarche inverse qui part des jugements moraux tels que nous

12. Cf. BARON & RITOV (1993), CARLSMITH, DARLEY & ROBINSON (2002), SUNSTEIN, SCHKADE & KAHNEMAN (2000).

les observons pour en déduire un mécanisme évolutionnaire. Comme nous allons le voir, une telle démarche conduit nécessairement à une approche mutualiste de la morale.

2.3 ⇨ Le contrat social: la morale comme équité

Les philosophes contractualistes (Rousseau, Kant, Rawls) ont depuis longtemps relevé la logique mutualiste des jugements moraux. C'est là le sens de l'analogie du contrat social: lorsque l'on cherche à se mettre d'accord avec les autres, il faut proposer un accord mutuellement avantageux. Selon l'approche mutualiste, agir moralement, c'est chercher l'équité, c'est-à-dire un *équilibre entre les intérêts de chacun* (pour une approche contemporaine du contractualisme, cf. KYMLICKA 1990, SCANLON 1998).

La logique de l'équilibre des intérêts n'est sans doute jamais aussi visible que dans un rapport de réciprocité. Observons la façon dont Lévi-Strauss décrit dans *Les Structures élémentaires de la parenté* l'offre réciproque de vin dans les restaurants bon marché. Dans ces établissements, il arrivait que l'on doive partager sa table avec un inconnu pour la durée du repas. Il était alors d'usage d'échanger son verre de vin.

Le partenaire, qui était en droit de se maintenir sur la réserve, est provoqué à en sortir; *le vin offert appelle le vin rendu, la cordialité exige la cordialité*. La relation d'indifférence, à partir du moment où l'un des convives décide d'y échapper, ne peut plus jamais se reconstituer telle qu'elle était; elle ne peut plus être, désormais que de cordialité ou d'hostilité: on n'a pas la possibilité, sans insolence, de refuser son verre à l'offre du voisin. Et l'acceptation de l'offre autorise une autre offre, celle de la conversation. Ainsi, toute une cascade de menus liens sociaux s'établissent, par une série d'oscillations alternées, selon lesquelles *on s'ouvre un droit en offrant, et on s'oblige en recevant*, et, dans les deux sens toujours, au delà de ce qui a été donné ou accepté. (LÉVI-STRAUSS 1949, p. 61.)

Le respect mutuel des intérêts individuels guide ainsi les interactions de la vie quotidienne et conduit à l'alternance de dons: le vin offert appelle le vin rendu, la conversation appelle la conversation en retour, le don exige un contre-don.

La réciprocité n'est qu'un exemple – le plus transparent sans doute – d'un rapport guidé par la logique de l'équilibre des intérêts. Dans leur étude sur la justice dans les rapports économiques, KAHNEMAN, KNETSCH & THALER (1986) montrent que cette logique se retrouve dans les situations les plus diverses.

9a. Une petite entreprise emploie plusieurs ouvriers et leur verse des salaires moyens; il y a un sérieux chômage dans la région et l'entreprise pourrait facilement remplacer ses employés par de bons ouvriers à des salaires plus bas. L'entreprise fait des bénéfices. La direction réduit les salaires de 5 %. Cela est-il acceptable? Sur 195 personnes, 23 % trouvent la réduction acceptable, 77 % la jugent inacceptable.

9b. Une petite entreprise [*idem*]. L'entreprise perd de l'argent. La direction réduit les salaires de 5 %. Cela est-il acceptable? Cette fois, les répondants jugent la réduction acceptable dans 68 % des cas et inacceptable dans 32 %.

Dans la situation 9a, l'entreprise ne vise pas le respect mutuel des intérêts de chacun. Elle cherche au contraire à prendre plus que sa part des bénéfices, en diminuant le salaire des salariés sans diminuer ses propres gains. Dans la situation 9b, en revanche, la diminution des salaires paraît plus légitime. Il semble que chacun doit accepter des sacrifices. Si l'entreprise perd de l'argent, les salariés doivent aussi accepter un salaire moins élevé.

5a. Une attente se développe pour un modèle populaire d'automobile. Les clients doivent désormais subir un délai de livraison de deux mois. Un concessionnaire vendait ses voitures au prix du catalogue. Il les propose désormais à un prix supérieur de 200 dollars au prix du catalogue. La décision est jugée acceptable pour 29 % des répondants, inacceptable par 71 % (N = 130).

5b. [...] Un concessionnaire vendait ses voitures en proposant un rabais de 200 dollars. Il les propose désormais au prix du catalogue. La réponse est jugée acceptable par 58 % et inacceptable par 42 % des répondants (N = 123).

Dans la situation 5a, le concessionnaire ne respecte pas les intérêts des clients. Il essaie au contraire d'obtenir le plus d'argent possible. Au contraire, dans la situation 5b, l'augmentation du prix des voitures est considérée comme légitime. Le concessionnaire ne fait que revenir sur un avantage qu'il avait concédé aux clients¹³.

13. Comme le note BOUDON (1995), les descriptions des situations proposées par Kahneman, Ketsch & Thaler laissent à l'interprétation une plus ou moins grande place. Il faut en effet compléter certains blancs (les actionnaires reçoivent-ils des dividendes? le concessionnaire est-il en situation de monopole? combien représente 200 dollars par rapport au prix d'une voiture?, etc.). Selon la façon dont on remplit ces blancs, on interprétera la situation de manière différente. C'est ce qui peut expliquer la plus ou moins

L'étude de Kahneman, Knetsch et Thaler illustre bien le fait que, pour être moraux, nos rapports doivent respecter mutuellement les intérêts de chacun. Les situations de justice distributive nous en offrent d'autres exemples. Ainsi, dans le monde du travail, la plupart des gens estiment que le salaire doit être en rapport avec la contribution (MARSHALL, SWIFT, ROUTH & BURGOYNE 1999). Si, en effet, j'ai contribué à l'entreprise commune dans une plus grande mesure que les autres, mes intérêts ne seront pas respectés si je reçois autant qu'eux. On m'aura en effet volé une partie de ce que j'ai apporté à l'entreprise commune. Mes intérêts ne seront respectés que si je reçois une part supérieure, en rapport avec l'importance de ma contribution. Ainsi, ceux qui apportent plus doivent recevoir plus, et ceux qui doivent recevoir moins sont aussi ceux qui ont apporté moins. Dans une enquête récente effectuée parmi des travailleurs français, 73 % des personnes interrogées pensent que l'écart de salaire entre les caissières et les cadres est trop important, alors qu'elles ne sont que 32 % à penser que l'écart de salaire entre les caissières et les médecins est excessif (DUBET 2006). Dans la mesure où les médecins ont fait de longues études et contribuent de manière très significative au bien-être des autres, il semble acceptable aux personnes interrogées qu'ils soient mieux payés que les caissières. En revanche, la rémunération des cadres ne semble pas reposer sur l'importance de leur contribution mais plutôt sur des privilèges. Bien sûr, l'équilibre des intérêts sera influencé par la façon dont chacun se représente les intérêts en jeu. On accordera un salaire moindre au médecin si l'on estime qu'ils ont poursuivi des études parce qu'ils venaient d'un milieu privilégié.

grande asymétrie entre les réponses. Encore une fois, une même logique peut conduire à des jugements différents, selon la façon dont on se représente la situation. La situation suivante laisse sans doute moins de place à l'interprétation : « 14. Un propriétaire loue une petite maison. Au moment du renouvellement du bail, il apprend que son locataire a pris un emploi tout près de son domicile. Il y a par conséquent peu de chances qu'il veuille déménager. Le propriétaire augmente alors son loyer de 40 % de plus qu'il ne l'avait envisagé. 9 % des participants jugent la décision acceptable ; 91 % la jugent inacceptable (N = 175). »

Dans cette situation, il y a peu de doutes que le propriétaire ne cherche pas à tenir compte des intérêts de son locataire et à lui proposer une relation mutuellement avantageuse.

L'aspect mutualiste de nos jugements moraux transparait également dans les cas de justice rétributive. Les gens estiment ainsi que la peine ne doit pas viser à prévenir le crime mais à compenser le déséquilibre entre le criminel et sa victime (DARLEY & PITTMAN 2003). La peine doit donc logiquement être proportionnée au crime (ROBINSON, KURZBAN & JONES 2006, SUNSTEIN, KAHNEMAN & SCHKADE 1998). Le débat sur la peine de mort n'est ainsi pas une question d'efficacité de la peine de mort, mais de justice : pour les adversaires de la peine de mort, celle-ci est injuste car elle disproportionnée ; pour les partisans, au contraire, la mort du meurtrier est la juste compensation de la mort qu'il a causée.

Au passage, la théorie mutualiste rend compte du caractère apparemment paradoxal de la justice rétributive. Premier paradoxe : pourquoi faire souffrir un meurtrier alors que cela ne rend pas la vie à sa victime ? Second paradoxe : pourquoi dit-on payer sa dette alors qu'en réalité, l'emprisonnement d'un meurtrier coûte très cher ? Ces paradoxes disparaissent si l'on considère que notre morale vise l'équilibre des intérêts. La souffrance du criminel diminue le déséquilibre généré par le crime.

Les philosophes du contrat social se sont concentrés sur les situations de justice (GAUTHIER 1986, RAWLS 1971). Autrement dit, ils ont proposé des théories du *sens de la justice*, plutôt que des théories du *sens de la morale*¹⁴. Les devoirs de solidarité respectent pourtant la même logique de l'équité. Ainsi, comme nous l'avons vu, nous avons le devoir d'aider les autres, mais ce devoir a ses limites. J'ai par exemple le devoir de donner mon sang pour sauver des blessés, mais je n'ai pas celui de donner ma vie, quand bien même mon sacrifice permettrait de sauver un grand nombre de vie (pour une approche empirique, cf. BARON & MILLER 2000). La logique des devoirs dépend des intérêts de chacun : plus les autres ont besoin d'aide et plus nous avons le devoir de les aider ; à l'inverse, plus les aider nous est coûteux, et moins nous en avons le devoir. Autrement dit, nos devoirs correspondent à un « juste milieu » entre l'égoïsme et l'altruisme, entre l'indifférence aux autres et le sacrifice de soi.

Jusqu'ici, nous nous sommes appuyés sur deux approches, l'approche naturaliste et l'approche mutualiste. Cependant, aucune de ces deux

14. Sur les limites des théories de la morale restreintes à la justice, cf. RAWLS (1971, p. 43), THOMSON (1971) et GILLIGAN (1982).

approches ne parvient à elle seule à expliquer les phénomènes moraux. D'un côté, l'approche naturaliste rend compte de la spécificité des phénomènes moraux, mais ne peut expliquer leur logique mutualiste. De l'autre, l'approche mutualiste décrit correctement la logique des jugements moraux, mais n'éclaire pas l'origine de cette logique. Je propose donc de réunir ces deux approches en une seule théorie naturaliste et mutualiste de la morale. Les humains seraient ainsi *naturellement* équipés d'une adaptation visant le *respect mutuel* des intérêts de chacun.

3 ⇨ Une théorie naturaliste et mutualiste de la morale

Avant de disposer d'une véritable théorie naturaliste et mutualiste de la morale, il convient de préciser un certain nombre de points. Il est d'abord nécessaire de clarifier les mécanismes évolutifs susceptibles d'avoir sélectionné un sens moral mutualiste. Il faut ensuite éclaircir le fonctionnement psychologique de ce sens moral. Enfin, la théorie se doit de montrer en quoi les analyses mutualistes des phénomènes moraux diffèrent des analyses proposées par les théories altruistes et continuistes.

3.1 ⇨ Les mécanismes évolutifs : pourquoi être moral ?

Comment le sens moral a-t-il pu émerger ? Partons du fait qu'il est avantageux pour les individus de s'entraider (pour chercher de la nourriture, se protéger, s'assister mutuellement pendant les périodes difficiles, etc.). Dans ce contexte, chaque individu cherche les meilleurs partenaires avec lesquels collaborer. On peut se représenter une telle situation sous la forme d'un marché de l'emploi ou d'un « marché de l'entraide » où chacun est à la fois offreur et demandeur d'entraide (sur la notion de marché en théorie de l'évolution, cf. NOË & HAMMERSTEIN 1995).

Sur ce marché, les individus recherchent les individus les plus forts, les plus intelligents, les plus compétents, mais également ceux qui partagent les coûts et les bénéfices de la manière la plus avantageuse. La disposition morale constitue un compromis entre égoïsme et altruisme. D'un côté, les individus égoïstes, qui font passer leurs intérêts avant ceux des autres, peineront à trouver des alliés. À l'inverse, des individus altruistes, faisant systématiquement passer leurs intérêts après ceux des autres, seront exploités par leurs partenaires. Le marché de l'entraide conduit donc naturellement à la sélection d'individus qui visent l'équilibre des intérêts ou le « juste milieu » comme disait Aristote :

La manière d'agir conforme à la justice, est celle qui tient le milieu entre faire soi-même ce qui est injuste [exploiter les autres], et le souffrir de la part des autres [être exploités par les autres]; car, dans le premier cas, on a plus, et, dans le second, on a moins qu'on ne devrait avoir. (*Éthique à Nicomaque*, V.)

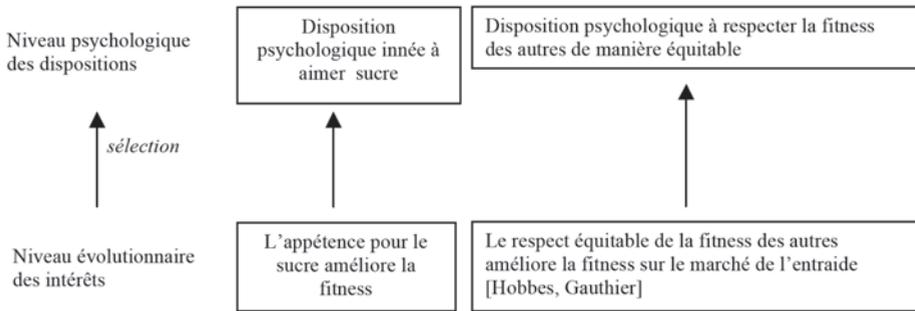
Le marché de la coopération donne donc un avantage aux individus «moraux», ceux qui, par exemple, partagent les produits de la chasse de manière équitable (en fonction de l'apport de chacun). Ils sont donc préférés aux autres et transmettent leurs dispositions morales à leurs descendants (pour une théorie similaire dans un cadre non naturaliste, cf. GAUTHIER 1986).

La nécessité de trouver des partenaires fiables pour collaborer peut paraître constituer une force trop faible pour faire émerger une disposition morale. Nul besoin de police, pourtant, entre les marchands internationaux du Moyen Âge ou entre les négociants chinois aujourd'hui : la crainte de perdre sa réputation est suffisante pour prévenir la tromperie (GREIF 1989, LANDA 1981). Le marchand qui vole un confrère gagne certes la valeur de la transaction mais ne trouve plus ensuite aucun partenaire susceptible lui proposer une offre. La réalité regorge d'exemples montrant que le seul intérêt à choisir des partenaires fiables conduit à l'exclusion des partenaires immoraux¹⁵. Selon la théorie mutualiste, le marché de l'entraide sélectionne donc progressivement et naturellement des individus dont le comportement vise spontanément à équilibrer les intérêts de chacun.

Si la théorie évolutionnaire que nous avons esquissé se vérifie, il y avait, *au niveau évolutionnaire*, un intérêt à se comporter moralement (cf. fig. 2). Les individus pourvus d'une disposition à respecter les intérêts de chacun étaient avantagés, ils étaient recherchés comme partenaires quand, au contraire, on se détournait des individus amoraux. *Au niveau psychologique*, les humains n'agissent pas, cependant, avec pour objectif d'améliorer leur position dans la société. Ils ne pensent pas non plus en terme d'avantage individuel.

Pour le dire autrement, on peut estimer que la théorie naturaliste et mutualiste s'accorde avec les théories contractariennes (Hobbes, Gauthier) au niveau évolutionnaire et avec les théories contractualistes

15. Pour une approche expérimentale, cf. CORICELLI, FEHR & FELLNER (2004), EHRHART & KESER (1999), PAGE, PUTTERMAN & UNEL (2005).



(Rousseau, Kant, Rawls, Scanlon) au niveau psychologique. Au niveau évolutionnaire, la théorie naturaliste considère, comme les théories contractariennes, que la morale existe parce qu'elle est avantageuse au niveau individuel. Au niveau psychologique, elle s'accorde avec les théories contractualistes (et contre les théories contractarienne) pour dire que les individus sont authentiquement moraux et qu'ils visent non pas l'avantage individuel mais l'équité.

3.2 ➔ Les mécanismes psychologiques : comment être moral ?

Jusqu'ici, j'ai uniquement parlé d'un mécanisme psychologique visant l'équilibre des intérêts. N'y a-t-il rien d'autre dans le sens moral qu'un tel mécanisme d'équilibre des intérêts ? Pour équilibrer les intérêts de chacun, il faut en effet être capable d'évaluer leurs conséquences au préalable, c'est-à-dire d'estimer si une action fait du bien ou du mal aux individus.

Il serait donc tentant d'intégrer à notre sens moral une *axiologie intuitive* (ou ce que l'on appelle en philosophie une théorie du bien), c'est-à-dire une capacité à évaluer ce qui affecte les intérêts individuels (les nôtres comme ceux d'autrui). Elle nous permet d'évaluer ce qui nous nuit et ce qui nous bénéficie en dehors de toute situation morale : nous évitons ainsi la douleur ou l'humiliation, nous aimons le confort et passer du temps en société. Ce sens des valeurs ou axiologie intuitive n'a cependant rien de spécifiquement moral. Pour veiller sur ses enfants, une mère doit savoir ce qui est bon pour eux. Pour torturer efficacement une victime, le tortionnaire doit savoir ce qui lui fait mal. L'évaluation de ce qui affecte autrui n'a donc rien de moral. En réalité, l'axiologie intuitive peut être constituée de la somme des dispositifs qui nous conduisent à accorder une valeur aux choses (l'affection pour nos enfants, notre souci de la réussite sociale, notre appétence pour

certaines nourriture, etc.). Le sens moral ne fait qu'utiliser ces intuitions pour juger si une action viole les intérêts d'autrui. Il faut donc distinguer le sens moral, qui vise le respect des intérêts de chacun, et l'axiologie intuitive, qui nous permet d'évaluer si une action est bonne ou non pour autrui.

Dans la mesure où le sens moral vise l'équilibre des intérêts individuels, tout ce qui est susceptible d'affecter les intérêts d'autrui peut être moralisé. Ces dernières années, la nourriture à elle seule a généré des débats moraux sur la taille des hamburgers, l'identité du producteur de café, la liberté des poulets, les kilomètres parcourus par une tomate, etc. (PINKER 2008). Voilà donc pourquoi un dispositif psychologique universel est compatible avec une très grande variété de jugements moraux. Ceux-ci dépendent des informations dont nous disposons concernant les coûts de chaque action pour autrui. Ainsi, on a longtemps considéré qu'il était acceptable de fumer en public. Fumer relevait en effet des préférences personnelles : certains n'aimaient pas fumer ou bien jugeaient que cela n'était pas bon pour leur santé. Ce n'est que lorsque l'on découvrit les méfaits du tabagisme passif que fumer en public devint réellement immoral. Tout est donc moralisable, pourvu qu'il apparaisse clairement que les intérêts d'autrui peuvent être affectés.

3.3 ➤ L'analyse mutualiste des phénomènes moraux

3.3.1 ➤ Les limites des théories continuistes et altruistes d'un point de vue mutualiste

Nous pouvons à présent mieux comprendre les limites des théories non mutualistes. Considérons d'abord la théorie continuiste. Un dispositif non moral comme la sympathie ne peut constituer la base des jugements moraux. En effet, toute souffrance n'est pas immorale. Ce qui est immoral, c'est la souffrance illégitime ou « indue » (PHARO 2004a), la souffrance qui résulte de l'injustice ou d'un manque de solidarité. Sympathie et sens moral ont deux logiques différentes. La sympathie est activée par la *peine* d'autrui, c'est-à-dire la souffrance quelle qu'elle soit. Le sens moral est activé par le *tort* fait à autrui, autrement dit la souffrance *illégitime*. L'analyse introspective du philosophe contractualiste Scanlon met en évidence le contraste entre les deux logiques de la sympathie et du sens moral.

Quand, par exemple, j'ai lu pour la première fois le fameux article de Singer sur la famine et senti la force de ses arguments, ce qui m'a convaincu, ce n'est n'était pas seulement *combien ces gens souffraient*

de la faim. Ce que j'éprouvais surtout, c'était le sentiment, différent, qu'il n'était pas moralement acceptable de ne pas les aider *étant donné la facilité avec laquelle je pouvais le faire*. (SCANLON 1998, p. 152, nous soulignons.)

Deux mécanismes sont activés dans l'exemple de Scanlon : la sympathie activée par la souffrance («*combien ces gens souffraient*») et le sens moral activé par le rapport entre le coût de l'aide et son bénéfice («*étant donné la facilité avec laquelle je pouvais le faire*»). Ces deux mécanismes sont indépendants. Nous pouvons avoir de la sympathie pour quelqu'un sans avoir le sentiment que nous devons lui venir en aide.

Il en va de même des rapports entre instinct parental et sens moral. Les parents n'ont pas des devoirs envers leurs enfants parce qu'ils les aiment (parce que leur instinct parental est activé). Ils ont des devoirs envers leurs enfants parce qu'ils sont donc responsables de la situation dans laquelle ceux-ci se trouvent et parce que leurs enfants ont besoin d'eux.

Enfin, le sens du dégoût ne dirige pas nos intuitions morales. Comme le reconnaît Haidt, le dégoût «éprouvé à l'encontre de déjections canines ou même d'un coït homosexuel, n'est pas un jugement moral en soi» (HAIDT 2007a). Le dégoût ne produit donc pas, par lui-même, de jugements moraux. Dans la culture occidentale, il est par exemple malpoli de roter à table et il est moralement inacceptable de cracher sur quelqu'un. Ces actions ne sont pas condamnées parce qu'elles impliquent le dégoût. Elles sont condamnées parce que l'on estime qu'elles nuisent à autrui et qu'il n'est pas très coûteux de s'abstenir de les commettre.

Nous comprenons mieux à présent pourquoi l'éthique des vertus ne parvient pas à expliquer pourquoi la sympathie est parfois une vertu (compassion) et parfois un vice (complaisance). Il lui manque la logique de l'équilibre des intérêts. Dans un cadre mutualiste, les dispositions amORALES comme la sympathie ne deviennent des vertus que lorsqu'elles permettent de réaliser ce que prescrit la logique du respect mutuel. On peut ainsi distinguer la sympathie de la compassion d'une part, et de la complaisance d'autre part. Avoir de la compassion, c'est éprouver de la sympathie pour ceux à qui l'ont *doit* de la sympathie, pour les personnes qui souffrent *injustement* ou pour celles qui sont *en droit* d'attendre notre assistance. Être complaisant, à l'inverse, c'est éprouver de la sympathie envers ceux qui ne le méritent pas. Un dispositif

psychologique peut être considéré comme une vertu lorsqu'il contribue à l'équilibre des intérêts individuels.

Nous comprenons mieux également les limites de la théorie altruiste. Reprenons les exemples étudiés dans la section 2. Dans les situations de justice distributive, la distribution ne vise pas le bien du groupe, parce qu'elle vise l'équilibre entre les intérêts de chacun (entre contribution et la rétribution). Dans les situations de justice rétributive, la punition ne vise pas le bien du groupe parce qu'elle vise le rétablissement de l'équilibre entre le criminel et sa victime. Enfin, nos devoirs de solidarité ne sont pas réglés par le bien du groupe (plus cela aide le groupe, plus nous avons le devoir d'aider) mais par l'équilibre des intérêts. C'est ce que montre encore l'analyse de Scanlon. Si nous avons des devoirs envers les affamés, ce n'est parce qu'ils souffrent, autrement dit parce que les soulager de leurs souffrances améliorerait considérablement le bien-être global, mais parce que cela coûte aussi peu à celui qui aide et que cela rapporte autant à celui qui est aidé.

3.3.2 ➔ Les dilemmes moraux dans une perspective mutualiste

Revenons, pour terminer, sur le dilemme du trolley. Pourquoi peut-on détourner le trolley vers le piéton seul, mais ne peut-on pas pousser le piéton se trouvant sur le pont? Pour le comprendre, considérons les positions respectives du piéton sur la voie (version standard) et du piéton sur le pont. Ces deux personnes ne sont pas dans la même position : en termes de sécurité, la position de l'homme sur le pont est en effet meilleure que celle de la personne sur la voie secondaire. En conséquence, en tuant le piéton qui se trouve en sécurité sur le pont, on lui cause un tort plus grand qu'en tuant la personne sur la voie secondaire.

Les différences de positions entre individus sont encore plus visibles dans les deux situations suivantes, adaptées de Thomson (1986).

(i) Six personnes sont en train de se noyer. Cinq se trouvent proches les unes des autres, la sixième se trouve au contraire loin du groupe. Une grosse bouée, reliée à une corde, a été lancée et flotte vers les six naufragés. La bouée se dirige vers la personne seule. Elle est encore très loin d'elle. Peut-on tirer sur la corde et diriger la bouée vers le groupe des cinq personnes?

(ii) Six personnes sont en train de se noyer... La bouée touche vers la personne seule. Celle-ci s'y accroche. Peut-on tirer sur la corde et récupérer la bouée pour la diriger vers les cinq personnes?

Dans la première situation, toutes les personnes ont à peu près les mêmes droits à la bouée. Dans la seconde, le fait que la personne seule se soit accrochée à la bouée lui donne davantage de droits à la conserver et donc à être sauvée : les six personnes ne sont donc plus dans la même position. L'une est sauvée, les autres sont toujours en train de se noyer. La position d'un individu est meilleure lorsqu'il touche la bouée que lorsqu'il en est encore loin. Par conséquent, on estime qu'en lui retirant la bouée, on lui cause un tort plus important. La logique de l'équilibre des intérêts accorde donc que l'on privilégie des cinq personnes dans la première condition (les six personnes sont à égalité), mais que l'on sauve en priorité la personne seule dans la seconde situation (pour une approche expérimentale, cf. BAUMARD, CHEVALIER & COVA, en préparation).

L'exemple de la bouée fait apparaître le lien entre le dilemme du trolley et une situation de justice distributive. La bouée est en effet un bien qui doit être distribué de la manière la plus juste. Dans le cas du trolley, en décidant qui doit mourir, on décide qui doit vivre et on distribue donc un bien : la vie. Comme dans une situation de justice, il y a conflit entre les intérêts des individus. Et comme dans une situation de justice, ce conflit doit être résolu en tenant compte des positions de chacun de manière à produire la distribution la plus équilibrée. Dans la version du pont piétonnier, il y a ainsi une grande différence entre la position du piéton et celle des cinq personnes sur la voie. Dans la version standard, au contraire, la position de la personne sur la voie n'est pas différente de celle des cinq autres. Comme le dit Judith THOMSON (1985), «ils sont sur un pied d'égalité». D'un point de vue intuitif, nous estimons que la personne seule n'a donc pas plus de droit à être sauvée que les cinq autres. Dans ces conditions, il est donc préférable de sauver cinq personnes. «On ne peut détourner le trolley que si la personne seule n'a pas plus de droits à éviter le trolley que les cinq personnes sur l'autre voie» (THOMSON 1985).

Plus généralement, la théorie mutualiste permet de rendre compte de l'existence même de dilemmes moraux. Dans un dilemme moral, la décision, quelle qu'elle soit, n'est pas conçue comme étant parfaitement satisfaisante. Au contraire, nous gardons le sentiment d'avoir pris la décision *la moins mauvaise*. C'est pour cela que l'on parle de dilemme. Dans le dilemme du trolley par exemple, les participants qui acceptent de détourner le trolley n'estiment pas, pour autant, avoir pris une *bonne décision*. L'existence de dilemmes moraux constitue notamment un problème pour la théorie altruiste (McCONNELL 2006). D'un point

de vue altruiste en effet, l'une des alternatives est toujours meilleure que les autres en terme de bien-être global¹⁶. Dans une perspective altruiste, la décision devrait ne jamais poser problème. Pourtant, ce n'est pourtant pas le cas.

La théorie mutualiste rend compte de l'intuition selon laquelle, dans certaines situations, la meilleure situation est malgré tout insatisfaisante. Dans le cas du trolley, comme dans celui de l'avortement, le bien en jeu (la vie) ne peut être partagé. Il faut choisir ou bien les cinq piétons, ou bien le piéton isolé, ou bien la mère, ou bien le fœtus. La logique de l'équilibre des intérêts ne peut donc s'appliquer jusqu'au bout. Quelque soit la solution, celle-ci ne sera pas équilibrée.

4 ↻ Conclusion: L'enjeu épistémologique d'une théorie naturaliste et mutualiste

Il reste évidemment beaucoup à faire pour disposer d'une véritable théorie naturaliste et mutualiste de la morale. Il est crucial en particulier de montrer comment la théorie rend compte de jugements moraux spécifiques (voir par exemple l'analyse mutualiste de l'avortement par THOMSON 1971). Il serait également nécessaire de préciser les mécanismes évolutionnaires et psychologiques. Il faudrait encore montrer dans quelle mesure cette théorie est valide à travers le temps et l'espace¹⁷. En dépit de tout cela, j'espère avoir montré l'intérêt d'une approche naturaliste et mutualiste.

Tout au long de cet article, nous nous sommes concentrés sur les théories de la morale. L'enjeu d'une théorie naturaliste et mutualiste dépasse cependant le simple champ des études de la morale et touche aux rapports entre sciences sociales et sciences naturelles. Une théorie

16. La théorie utilitariste admet certes des situations d'indécision. Il ne s'agit cependant pas de véritables dilemmes mais plutôt d'ignorance de certains paramètres qui nous empêche d'évaluer avec assurance les conséquences de nos actions. À la différence des dilemmes moraux, une information plus complète ferait disparaître cette situation. La théorie utilitariste ne rend donc pas compte de l'existence des dilemmes moraux.

17. Il existe déjà un certain nombre de travaux empiriques sur la validité inter-culturelle d'une approche mutualiste (cf. notamment NEFF 1997, TURIEL 2002). On peut également relire le débat entre MALINOWSKI (1926) et DURKHEIM (1893) sur les droits et les devoirs dans les sociétés apparemment collectivistes.

naturaliste présente un exemple possible d'une théorie intégrant sciences naturelles et sciences sociales.

4.1 ⇨ L'approche naturaliste : une approche non réductionniste

Historiquement, les sciences sociales ont cherché à obtenir leur autonomie institutionnelle et scientifique en se construisant contre les sciences naturelles (cf. par exemple GUILLO 2000). L'opposition aux sciences naturelles, et en particulier aux approches évolutionnaires, reposait également sur de bonnes raisons politiques, la théorie de l'évolution ayant trop souvent servi à légitimer des programmes ouvertement racistes, sexistes ou eugénistes.

Ces deux raisons ont aujourd'hui pratiquement disparues. Les approches évolutionnaires ont complètement abandonné le concept de race, et s'intéressent essentiellement aux capacités communes à tous les membres de l'espèce humaine (psychologie intuitive, gestion de la réputation, etc.). Les sciences sociales ont quant à elles acquis une autonomie institutionnelle et scientifique que personne ne songe à remettre en cause.

Surtout, les approches naturalistes contemporaines auxquelles ce travail entend contribuer ne cherchent pas à réduire les sciences sociales aux sciences naturelles, les phénomènes sociaux aux neurones ou aux gènes. Comme nous l'avons vu tout au long de ce travail, une approche naturaliste du sens moral ne fait pas disparaître le niveau psychologique au profit, par exemple, du niveau évolutionnaire. Si je m'abstiens de voler alors que je le pourrais, ce n'est pas parce que mes gènes ont intérêt à ce que je le fasse. C'est parce que je suis disposé naturellement à me comporter moralement et que, dans le contexte dans lequel je me trouve (un petit magasin), mon sens moral se trouve plus activé que mes désirs d'enrichissement. Mon comportement s'explique donc essentiellement par la conjonction de causes environnementales et de causes psychologiques. Nous pouvons aller plus loin. Au niveau sociologique, nous chercherons à rendre compte des causes environnementales (présence de témoins, représentations sociales concernant les propriétaires pauvres de petits magasins, etc.). Au niveau évolutionnaire, nous tenterons d'expliquer notre disposition morale en termes évolutionnaires (qu'est-ce qui a conduit à la sélection d'un sens moral?). À chacun son niveau d'explication : la psychologie rend compte du fonctionnement du sens moral ; la sociologie du contexte social qui l'active ; la théorie de l'évolution des causes qui l'ont sélectionné (SPERBER 1997a).

Une approche naturaliste ne cherche donc pas à réduire les sciences sociales à la biologie. Elle ne fait qu'offrir aux sciences sociales un ensemble d'outils susceptibles d'être utilisés dans l'explication des phénomènes sociaux (BAUMARD, CHEVALLIER & COVA, en préparation). Le comportement moral intervient ainsi dans de nombreux phénomènes sociaux. La capacité morale est alors un «mécanisme» (ELSTER 1998) parmi d'autres, comme le souci de sa réputation, la faiblesse de la volonté, le raisonnement, etc. À ce titre, elle s'intègre donc sans difficulté dans les explications causales des phénomènes sociaux (pour un modèle du genre, cf. VEYNE 1976).

4.2 ➔ L'intérêt d'une approche naturaliste

A-t-on besoin cependant d'une approche naturaliste de la morale ? Ne faut-il pas plutôt, comme le suggère BOUDON (1995, p. 446), «s'en tenir à la tradition aristotélicienne, pour laquelle le sens de la justice est une donnée, qu'il s'agit de décrire de façon aussi précise que possible»? Cette position non naturaliste semble justifiée. Comme je l'ai dit en introduction, les théories naturalistes ont le plus souvent été construites dans l'ignorance des théories descriptives proposées par les philosophes. Les naturalistes cherchaient essentiellement à montrer que la morale était naturelle. Les approches naturalistes n'éclairaient donc pas la logique contractualiste observée sur le terrain. Elles ne faisaient qu'intégrer la morale à un cadre plus général, sans fournir d'outils supplémentaires à l'analyse sociologique¹⁸. J'espère cependant avoir montré que d'autres approches naturalistes sont possibles. Elles partent des jugements moraux, et les éclairent en retour à l'aide de la psychologie et de la théorie de l'évolution.

Notons tout d'abord que la question de la coopération est une question centrale des sciences sociales (NISBET 1966). Or, depuis le début, les sciences sociales, et en particulier la sociologie, butent sur ce que KANT (1999 [1784], p. 15) appelait l'«insociable sociabilité de l'homme». D'un côté, l'homme a une inclination à s'associer, à sympathiser et à respecter les intérêts d'autrui. De l'autre, il semble avant tout gouverné par ses propres intérêts. Il y a là un paradoxe, comme le reconnaît ROUSSEAU

18. De fait, Boudon fait ici référence au livre de Robert WILSON, *Le Sens moral* (1993), qui cherche à montrer le caractère naturel de la morale sans s'attacher à défendre un mécanisme en particulier.

(2002 [1762], p. 72) : «Se préférer à tout est un penchant naturel à l'homme, écrit-il dans *Émile*, pourtant le premier sentiment de la justice est inné dans le cœur humain.»

Les solutions à ce paradoxe sont souvent peu satisfaisantes. Certaines, à la suite de Hobbes, proposent de ramener la sociabilité humaine à de l'égoïsme bien compris. Ces solutions possèdent une part de vérité. Nous respectons parfois notre devoir pour nous mettre en avant ou pour obliger les autres à nous rendre service. De plus, notre comportement moral, s'il ne vise pas directement notre intérêt personnel, nous est bien souvent utile. Cependant, comme l'illustre l'anecdote de Mencius, ces solutions ne rendent pas compte de nos actes authentiquement moraux. À l'inverse, d'autres solutions consistent à expliquer la morale en terme d'altruisme. Il arrive en effet aux humains d'agir de manière altruiste, en particulier à l'égard de leurs enfants. Le domaine de l'altruisme est cependant limité (parents, enfants, amis). Comme nous l'avons vu tout au long de ce travail, nos comportements moraux, en particulier, ne relèvent pas de l'altruisme.

Une approche naturaliste et mutualiste permet de sortir de cette opposition. Au niveau des motivations psychologiques, le comportement moral ne relève ni de l'égoïsme, ni de l'altruisme. Il vise au contraire le respect mutuel des intérêts de chacun. Au niveau évolutionnaire, le comportement moral n'est pas en contradiction avec l'égoïsme puisque, dans la société humaine, il est dans notre intérêt de respecter les intérêts d'autrui. Nous disposons donc d'une façon de concilier l'intérêt individuel avec l'existence de motivations authentiquement morale. Rappelons-nous les mots de RAWLS (1971) : la société humaine se caractérise à la fois par un conflit d'intérêts et par une identité d'intérêts. La morale vise précisément à concilier l'intérêt individuel et la nécessité de s'associer. L'«insociable sociabilité» de l'homme n'a donc plus rien de paradoxal. D'une part, la morale est dans l'intérêt des individus. D'autre part, elle obéit à la logique de l'avantage mutuel.

Comme je l'ai noté à plusieurs reprises, la solution mutualiste n'est pas neuve. Les philosophes contractualistes l'ont défendu depuis longtemps. Ils ont notamment fourni des *descriptions* détaillées de notre capacité morale (RAWLS 1971, THOMSON 1971). Cependant, ils n'ont jamais pu *expliquer* pourquoi les humains sont pourvus d'une telle capacité (pourquoi nos jugements visent-ils l'avantage mutuel ? pourquoi nous comportons-nous moralement ?). Or, sans explication, la théorie

mutualiste paraît improbable : pourquoi agirions-nous comme si nous avions passé un contrat alors que, de toute évidence, ce n'est pas le cas ? De fait, les descriptions des philosophes contractualistes ont rarement été intégrées aux enquêtes empiriques (cf. cependant DUBET 2006, ELSTER 1992).

La position non naturaliste n'est donc pas satisfaisante. Se contenter de décrire notre capacité morale n'est pas suffisant. L'absence d'une explication du sens moral affaiblit nécessairement la description qu'en font les philosophes contractualistes. Il est toujours tentant en effet de ramener la morale à autre chose, soit à la manipulation et à la fausse conscience, soit à l'altruisme ou aux normes sociales¹⁹. On identifie alors la morale au sacrifice ou à l'égoïsme, et l'on oublie les descriptions des philosophes contractualistes.

Une approche naturaliste fournit également une meilleure description de notre capacité morale, de la façon dont elle s'active, de son fonctionnement et des jugements qu'elle produit. Cette description est cruciale pour l'enquête empirique. Premièrement, il est souvent difficile de s'assurer dans un contexte donné des motivations réelles des individus (cf. notamment ELSTER 2007, p. 95-111) : essaient-ils de manipuler les autres ? Agissent-ils de manière authentiquement morale ? Mieux connaître le sens moral (et notamment les émotions morales) permet

19. Boudon n'échappe pas à la tentation. Constatant que la plupart des jugements moraux correspondent à la logique contractualiste (BOUDON 1995, p. 352), il propose un éclairage d'inspiration fonctionnaliste. « Pourquoi y aurait-il "démotivation" des acteurs à partir du moment où les différences de rémunération s'écarterait de l'optimum ? Parce qu'un sentiment d'injustice se ferait jour, les acteurs ayant alors l'impression que le parallélisme entre rétribution et contribution est brisé. Pourquoi ce parallélisme est-il perçu comme important ? Parce qu'il est constitutif du système lui-même. Les employés sont sensés contribuer à la finalité du système ; c'est la raison même de leur participation ; quant à leur salaire, il rémunère cette contribution ; il faut donc qu'il en reflète l'importance. C'est pourquoi la violation du principe de l'égalité entre rétribution et contribution n'est pas tolérable : elle est contradictoire avec les règles de base qui émanent de la nature même du système » (*ibid.*, p. 271).

Une telle théorie n'a rien d'explicatif. Ce n'est pas parce que nous avons tous intérêt à ne pas aggraver l'effet de serre que nous réduisons effectivement nos rejets de gaz carbonique.

donc d'identifier avec plus de sûreté les comportements moraux. La théorie naturaliste permet en outre de mieux distinguer le sens moral d'autres dispositions comme la sympathie, le souci de la réputation ou l'affection envers ses proches.

Deuxièmement, en clarifiant les mécanismes impliqués dans le jugement moral (axiologie intuitive, effets de cadrage, domaine effectif, microdispositions), une théorie naturaliste du sens moral améliore notre compréhension des jugements observés sur le terrain. Elle fournit une grille de lecture (évaluation des intérêts, équilibre des intérêts) des comportements et des jugements individuels.

À un niveau plus théorique, l'approche naturaliste apporte des contraintes aux explications des phénomènes sociaux. Si le sens moral vise bien le respect mutuel des intérêts individuels, dès lors que l'on qualifie un jugement de moral, nous devons chercher à en rendre compte dans les termes de la théorie. Il faut montrer qu'une action est condamnée parce que les gens jugent qu'elle affecte les intérêts individuels, et qu'elle ne respecte pas également les intérêts de toutes les parties en jeu. Bien souvent, pourtant, les enquêtes empiriques se contentent d'enregistrer l'existence d'un jugement moral et d'en rendre compte à l'aide d'une norme culturelle *ad hoc*.

Examinons par exemple le cas de l'usure, la pratique qui consiste à prêter de l'argent avec intérêt. Elle fut condamnée au Moyen Âge avant de devenir respectable avec le développement des investissements commerciaux et industriels. Il serait aisé d'expliquer ce phénomène en invoquant un « changement des valeurs » (la société serait ainsi passée de valeurs chrétiennes aux valeurs capitalistes). Une telle explication n'explique cependant pas grand chose. Elle ne fait que reformuler le constat du passage de la condamnation de l'usure à son autorisation. En réalité, cette différence de jugement tient non pas à un changement dans les principes moraux mais plutôt à l'évolution de la pratique de l'usure. Ainsi, au Moyen Âge, les prêts étaient utilisés principalement pour la consommation, alors qu'aujourd'hui ce sont des investissements dans des entreprises capitalistes. La différence est très pertinente. Dans le premier cas, l'usurier profite de la pauvreté de celui à qui il prête. Dans le second cas, l'usurier demande une juste part des bénéfices de l'entreprise dans laquelle il a investi (LUKES 2008). Les mêmes intuitions conduisent à condamner l'usure dans le premier cas et à l'accepter dans le second cas.

Un tel exemple montre qu'une théorie naturaliste de la morale ne vise pas à remplacer l'analyse sociologique, mais bien à la seconder en lui fournissant des outils supplémentaires. Une théorie naturaliste de la morale ne se substitue donc pas aux analyses sociologiques (PHARO 2004b). Elle n'explique rien par elle-même. Elle nécessite au contraire un contexte social et culturel pour pouvoir fonctionner.

Plus généralement, les théories en sciences sociales souffrent de ne pas être des théories causales (ELSTER 2007, p. 7-67, SPERBER 1997a). Une théorie causale est une théorie qui explique un phénomène à l'aide d'un enchaînement de mécanismes causaux, eux-mêmes explicables en termes causaux plus élémentaires.

Les théories du choix rationnel ne sont pas en elles-mêmes des théories causales. En effet, montrer qu'une action est rationnelle ne revient pas à l'expliquer, c'est-à-dire à fournir la chaîne causale qui a conduit à son existence. On peut par exemple montrer qu'il est rationnel, dans une situation donnée, de ne pas voler l'argent du magasin (parce qu'il y a une caméra de surveillance, parce que je me trouve avec des amis qui auront ensuite des doutes sur mon honnêteté, etc.). Pour autant, en montrant qu'il est rationnel, nous n'expliquons pas mon comportement: ai-je agi pour préserver ma réputation? Par peur de la police? Pour des raisons authentiquement morales? Pour que la théorie du choix rationnel soit valide, il faudrait montrer à chaque fois que mon comportement découle bien d'un raisonnement effectif sur la situation, et que je suis mû par ce raisonnement. Au vu des travaux empiriques que nous avons passé en revue, il est vraisemblable de penser que ce cas de figure se présente rarement.

Il en va de même des explications en termes de culture ou de normes sociales. Certes, dans la société dans laquelle je me trouve, il est mal vu de voler dans la caisse des magasins. Or, je m'abstiens de voler. Mon comportement est-il causé par la culture de ma société? Et d'abord, comment est-il causé? Comment la culture agit-elle sur mon cerveau? Une telle théorie ne nous fournit aucun mécanisme causal à même d'expliquer mon comportement. Parler d'intériorisation, ce n'est que donner un label à un phénomène, et non l'expliquer. La théorie holiste ne rend pas compte non plus du fait que certaines normes culturelles semblent parfois illégitimes aux membres d'une société. Pourquoi certaines normes agissent-elles sur notre comportement alors que d'autres rencontrent beaucoup de résistance? Bien des théories holistes sont

tentées d'invoquer la fonctionnalité des normes morales. La morale semble en effet contribuer au bien commun, elle contraint les individus et limite leur égoïsme. Or de telles théories fonctionnalistes ne sont pas en elles-mêmes causales. Elles n'expliquent pas pourquoi les normes d'une société sont fonctionnelles. Les sociétés ne sont pas construites par des Grands Ingénieurs, elles ne sont pas sélectionnées par la sélection naturelle ni par aucun mécanisme de sélection évident. En réalité, il y a apparence de fonctionnalité. Les normes morales contribuent bel et bien au bien commun, mais elles ne sont qu'un sous-produit d'une disposition avantageuse au niveau individuel. De fait, de la malnutrition aux désastres écologiques, les exemples de dysfonctionnement des normes sociales ne manquent pas.

La rareté des explications causales en sciences sociales tient sans doute à ce que les mécanismes élémentaires susceptibles d'être utilisés demeurent mal connus. C'est là tout l'intérêt d'une approche naturaliste. Comme nous l'avons vu dans le cas de la morale, l'approche naturaliste permet d'expliquer la raison d'être de ces mécanismes et de les décrire de manière elle-même mécaniste. Mieux ces mécanismes seront connus, et plus il sera aisé et fécond d'expliquer causalement les phénomènes sociaux (ELSTER 2007). Montrer que l'approche naturaliste est non seulement compatible avec les sciences sociales, mais qu'elle leur est également hautement utile : voilà tout l'enjeu d'une théorie naturaliste et mutualiste de la morale.

☉Remerciements☉

Je remercie Florian Cova, Coralie Chevallier, Michel Forsé et Dan Sperber pour leurs commentaires sur les versions précédentes de cet article.

☉Références bibliographiques☉

A

ALESINA A. & GLAESER E. (2004), *Fighting poverty in the US and Europe: A world of difference*, Oxford University Press.

B

BARON J. (1995), "Blind justice: fairness to groups and the do-no-harm principle", *Journal of Behavioral Decision Making*, 8(2): 71-83.

BARON J. & MILLER J. (2000), "Limiting the Scope of Moral Obligations to Help: A Cross-Cultural Investigation", *Journal of Cross-Cultural Psychology*, 31(6): 703.

- BARON J. & RITOV I. (1993), "Intuitions about penalties and compensation in the context of tort law", *Journal of Risk and Uncertainty*, 7(1): 17-33.
- BAUMARD N. (2007), "La morale n'est pas le social. Le point de vue de la psychologie", *Terrain*, 48: 49-72.
- BAUMARD N. (2008), *Une théorie naturaliste et mutualiste de la morale*, thèse de doctorat à l'École des hautes études en sciences sociales, Paris.
- BAUMARD N., CHEVALLIER C. & COVA F. (en préparation), "A mutualistic approach to moral dilemmas".
- BAUMARD N., CHEVALLIER C. & MASCARO O. (soumis), "It's not fair - The sense of justice in young children".
- BAUMARD N. & SPERBER D. (2007), "La morale", *Terrain*, 48: 5-12.
- BOUDON R. (1995), *Le Juste et le vrai. études sur l'objectivité des valeurs et de la connaissance*, Fayard.
- BOYD R., GINTIS H., BOWLES S. & RICHESON P. (2003), "The evolution of altruistic punishment", *Proc. Natl. Acad. Sci. USA*, 100(6): 3531-3535.
- BUTLER J. (2008), *Fifteen Sermons on Human Nature* [1726], Project Gutenberg (www.gutenberg.org/wiki/Main_Page).

C

- CALLEBAUT W. & RASSKIN-GUTMAN D. (2005), *Modularity: Understanding The Development And Evolution of Natural Complex Systems*, MIT Press.
- CARLSMITH K., DARLEY J. & ROBINSON P. (2002), "Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment", *Journal of Personality and Social Psychology*, 83(2): 284-299.
- CORICELLI G., FEHR D. & FELLNER G. (2004), "Partner Selection in Public Goods Experiments", *Journal of Conflict Resolution*, 48(3): 356-378.

D

- DAMON W. (1988), *The Moral child: Nurturing children's natural moral growth*, Free Press.
- DARLEY J. & PITTMAN T. (2003), "The Psychology of Compensatory and Retributive Justice", *Personality and Social Psychology Review*, 7(4): 324-336.
- Darwin C. (1859), *On the Origin of Species by Means of Natural Selection*, J. Murray.
- Darwin C. (1871). *The Descent of Man, and Selection in Relation to Sex*, D. Appleton and Company. Trad. fr., *La Descendance de l'homme et la sélection sexuelle*, Reinwald et Cie, 1891.
- DE WAAL F. (1996), *Good natured: The origins of right and wrong in humans and other animals*, Harvard University Press.
- DEHAENE S. (1997), *La Bosse des maths*, Odile Jacob.
- DEHAENE S. (2008), *Les Neurones de la lecture*, Odile Jacob.
- DUBET F. (2006), *Injustices: l'expérience des inégalités au travail*, Seuil.
- DUNN J. (1988), *The Beginnings of social understanding*, Harvard University Press.
- DURKHEIM É. (1893), *De la division du travail social*, F. Alcan.

E

- EHRHART K.-M. & KESER C. (1999), "Mobility and Cooperation: On the Run", *Série scientifique* (CIRANO): 99s-24.
- ELSTER J. (1992), *Local justice: how institutions allocate scarce goods and necessary burdens*, Russell Sage Foundation.
- ELSTER J. (1998), "A plea for mechanisms" (p. 45-73), in P. Hedström & R. Swedberg (eds.), *Social mechanisms: an analytical approach to social theory*, Cambridge University Press.

ELSTER J. (2007), *Explaining social behavior: More nuts and bolts for the social sciences*, Cambridge University Press.

ELSTER J. (2009), *Le Désintéressement. Critique de l'homme économique*, Le Seuil.

EVANS J. (2008), "Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition", *Annual Review of Psychology*, 59.

F

FOOT P. (1967), "The problem of abortion and the doctrine of double effect", *Oxford Review*, 5: 5-15.

G

GAUTHIER D. (1986), *Morals by agreement*, Clarendon Press, Oxford University Press.

GIBBARD A. (1990), "Norms, Discussion, and Ritual: Evolutionary Puzzles", *Ethics*, 100(4): 787-802.

GREIF A. (1989), "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders", *Journal of Economic History*, 49(4): 857-882.

GUILLO D. (2000), *Sciences sociales et sciences de la vie*, PUF.

H

HAIDT J. (2001), "The emotional dog and its rational tail: A social intuitionist approach to moral judgment", *Psychological Review*, 108: 814-834.

HAIDT J. (2007a), "Doing science as if groups existed", <www.edge.com>.

HAIDT J. (2007b), "The New Synthesis in Moral Psychology", *Science*, 316(5827): 998.

HAIDT J. & HERSH M. (2001), "Sexual morality: The cultures and emotions of conservatives and liberals", *Journal of Applied Social Psychology*, 31(1): 191-221.

HAIDT J. & JOSEPH C. (2004), "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues", *Daedalus* (Special issue on human nature): 55-66.

HAUSER M., CUSHMAN F., YOUNG L. & JIN R. (2007), "A dissociation between moral judgments and justifications", *Mind & Language*, 22(1): 1-21.

HEYD D. (1982), *Supererogation: Its status in ethical theory*, Cambridge University Press.

HINDE R. (2002), *Why good is good: The sources of morality*, Routledge.

HURSTHOUSE R. (2007), "Virtue Ethics", in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, Stanford University.

HUTCHESON F. (2004), *An inquiry into the original of our ideas of beauty and virtue: in two treatises* [1725], Liberty Fund.

K

KAGAN S. (1989), *The Limits of morality*, Clarendon Press, Oxford University Press.

KAHN P.H.J. (1992), "Children's Obligatory and Discretionary Moral Judgments", *Child Development*, 63(2): 416-430.

KAHNEMAN D., KNETSCH J. & THALER R. (1986), "Fairness as a Constraint on Profit Seeking: Entitlements in the Market", *The American Economic Review*, 76(4): 728-741.

KANT E. (1999), *Idée d'une histoire universelle d'un point de vue cosmopolitique* [1784], Bordas.

KETELAAR T. & AU W.T. (2003), "The effects of guilty feelings on the behavior of uncooperative individuals in repeated social bargaining games: An Affect-as-information interpretation of the role of emotion in social interaction", *Cognition & Emotion*, 17: 429-453.

KOHLBERG L. (1969), *Stage and Sequence: The Cognitive-developmental Approach to Socialization*, Rand McNally.

KROPOTKINE P. (2001), *L'Entraide: un facteur de l'évolution* [1902], Écosociété.

KUMAR R. (1999), "Defending the Moral Moderate: Contractualism and Common Sense", *Philosophy and Public Affairs*, 28(4): 275-309.

KYMLICKA W. (1990), *Contemporary political philosophy: An introduction*, Oxford University Press. Trad.fr., *Les Théories de la justice: une introduction*, La Découverte.

L

LANDA J.T. (1981), "A Theory of the Ethnically Homogeneous Middleman Group: An Institutional Alternative to Contract Law", *The Journal of Legal Studies*, 10(2): 349-362.

LESLIE A.M., MALLON R. & DICORCIA J.A. (2006), "Transgressors, victims, and cry babies: Is basic moral judgment spared in autism?", *Social Neuroscience*, 1(3 & 4): 270-283.

LÉVI-STRAUSS C. (1949), *Les Structures élémentaires de la parenté*, PUF.

LUKES S. (2008), *Moral Relativism*, Picador.

M

MACINTYRE A.C. (1981), *After Virtue: A Study in Moral Theory*, Gerald Duckworth & Co Ltd.

MALINOWSKI B. (1926), *Crime and custom in savage society*, Harcourt, Brace & Company.

MARSHALL G., SWIFT A., ROUTH D. & BURGOYNE C. (1999), "What Is and What Ought to Be: Popular Beliefs about Distributive Justice in Thirteen Countries", *European Sociological Review*, 15(4): 349-367.

MCCONNELL T. (2006), *Moral Dilemmas*, Stanford University Press.

MICHELBACH P.A., SCOTT J.T., MATLAND R.E. & BORNSTEIN B.H. (2003), "Doing Rawls Justice: An Experimental Study of Income Distribution Norms", *American Journal of Political Science*, 47(3): 523.

MIKHAIL J., SORRENTINO C.M. & SPELKE E. (1998), "Toward a Universal Moral Grammar", paper presented at the Twentieth Annual Conference of the Cognitive Science Society.

MITCHELL G., TETLOCK P., NEWMAN D.G. & LERNER J.S. (2003), "Experiments Behind the Veil: Structural Influences on Judgments of Social Justice", *Political Psychology*, 24: 519.

MITCHELL G., TETLOCK P.E., MELLERS B.A. & ORDÓÑEZ L.D. (1993), "Judgments of social justice: Compromises between equality and efficiency", *Journal of Personality and Social Psychology*, 65: 629-639.

N

NEFF K. (1997), *Reasoning about Rights and Duties in the Context of Indian Family Life*, University of California.

NICHOLS S. (2004), *Sentimental Rules: On the Natural Foundations of Moral Judgment*, Oxford University Press.

NISBET R.A. (1966), *The Sociological Tradition*, Basic Books.

NOE R. & HAMMERSTEIN P. (1995), "Biological markets", *Trends in Ecology and Evolution*, 10(8): 336-339.

O

OHMAN A., FLYKT A. & ESTEVES F. (2001), "Emotion Drives Attention: Detecting the Snake in the Grass", *Journal of Experimental Psychology (General)*, 130(3): 466-478.

P

PAGE T., PUTTERMAN L. & UNEL B. (2005), "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry and Efficiency", *The Economic Journal*, 115(506): 1032-1053.

PHARO P. (2004a), «L'enquête en sociologie morale», *L'Année sociologique* (1949), 54(2): 359-388.

PHARO P. (2004b), *Morale et sociologie*, Gallimard.

PIAGET J. (1932), *Le Jugement moral chez l'enfant*, PUF.

PINKER S. (1997), *How the mind works*, Norton.

R

RAWLS J. (1971), *A Theory of Justice*, Belknap Press of Harvard University Press. Trad. fr., *Théorie de la justice*, Seuil, 1997.

RIDLEY M. (1996), *The Origins of Virtue*, Viking.

ROBINSON P.H., KURZBAN R. & JONES O.D. (2006), "The Origins of Shared Intuitions of Justice", University of Pennsylvania Law School, Public Law Working Paper, No. 06-47.

ROSS H. (1996), "Negotiating principles of entitlement in sibling property disputes", *Developmental Psychology*, 32(1): 90-101.

ROUSSEAU J.-J. (2002), *Émile* (vol. IV) [1762], Les classiques des sciences sociales (<<http://classiques.uqac.ca/>>).

ROZIN P., HAIJT J. & MCCAULEY C. (1993), "Disgust", in M.L.J. Haviland (ed.), *Handbook of emotions*, Guilford Press: 575-594.

S

SCANLON T. (1998), *What we owe to each other*, Belknap Press of Harvard University Press.

SCHOKKAERT E. (1999), «M. Tout-le-monde est "post-welfariste": Opinions sur la justice redistributive», *Revue économique*, 50(4): 811-831.

SCHOLL B.J. & LESLIE A.M. (1999), "Modularity, Development and "Theory of Mind"", *Mind and Language*, 14(1): 131-153.

SHAFTESBURY A.A.C. (1714), *Characteristicks of men, manners, opinions, times*, 2d ed., London.

SMITH A. (1999), *Théorie des sentiments moraux* [1759], PUF.

SOBER E. & WILSON D. (1998), *Unto others: the evolution and psychology of unselfish behavior*, Harvard University Press.

SPERBER D. (1993), «Quelques remarques sur le relativisme moral d'un point de vue anthropologique», in J.-P. Changeux (dir.), *Les Fondements naturels de l'éthique*, Odile Jacob.

SPERBER D. (1997a), «Individualisme méthodologique et cognitivisme» (p. 123-136), in R. Boudon, F. Chazel & A. Bouvier (dir.), *Cognition et sciences sociales*, PUF.

SPERBER D. (1997b), "Intuitive and reflective beliefs", *Mind and Language*, 12 (1): 67-83.

SPERBER D. & HIRSCHFELD L. (2004), "The cognitive foundations of cultural stability and diversity", *Trends in Cognitive Sciences*, 8(1): 40-46.

SUNSTEIN C., KAHNEMAN D. & SCHKADE D. (1998), "Assessing Punitive Damages (With Notes on Cognition and Valuation in Law)", *Yale Law Journal*, 107(7): 2071-2153.

SUNSTEIN C., SCHKADE D. & KAHNEMAN D. (2000), "Do People Want Optimal Deterrence?", *Journal of Legal Studies*, 29(1): 237-253.

T

TAKAHASHI H., YAHATA N., KOEDA M., MATSUDA T., ASAI K. & OKUBO Y. (2004), "Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study", *Neuroimage*, 23(3): 967-974.

TANGNEY J.P. & DEARING R.L. (2002), *Shame and Guilt*, The Guilford Press.

THOMSON J.J. (1971), "A Defense of Abortion", *Philosophy & Public Affairs*, 1(1): 47-66.

THOMSON J.J. (1985), "The Trolley Problem", *Yale Law Journal*, 94(6): 1395-1415.

THOMSON J.J. (1986), *Rights, restitution, and risk: essays in moral theory*, Harvard University Press.

TISAK M.S. & TURIEL E. (1984), "Children's Conceptions of Moral and Prudential Rules", *Child Development*, 55(3): 1030-1039.

TURIEL E. (1983), *The Development of social knowledge: morality and convention*, Cambridge University Press.

TURIEL E. (2002), *The Culture of morality: social development and social opposition*, Cambridge University Press.

TYLER T. & WEBER R. (1983), "Support for the death penalty", *Law and Society Review*, 17: 201-224.

V

VEYNE P. (1976), *Le Pain et le cirque. Sociologie historique d'un pluralisme politique*, Seuil.

W

WEBER M. (1995), *Économie et société* [1921], Pocket.

WILSON E.O. (1975), *Sociobiology: The New Synthesis*, Harvard University Press.

WILSON J. (1993), *The Moral Sense*, Free Press.

☞ Notice biographique ☞

Nicolas Baumard est chercheur à l'Institut d'anthropologie évolutionnaire et cognitive de l'Université d'Oxford. Après des études en sciences naturelles (biologie, psychologie) et en sciences humaines (sociologie, philosophie), il a effectué sa thèse sous la direction de Dan Sperber sur les origines de la morale. Il a développé une théorie naturaliste et contractualiste de la morale à travers de nombreux articles et d'un livre à paraître en octobre 2010 aux éditions Odile Jacob.

Avertissement. Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur et des auteurs. Les textes et illustrations figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage personnel, scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document selon ce modèle :

Nicolas Baumard, « Une théorie naturaliste des phénomènes moraux est-elle possible ? », in Gérard Lambert et Marc Silberstein (dir.), Matière première. Revue d'épistémologie [en ligne]. Nouvelle série, N° 1/2010 : Épistémologie de la médecine et de la santé, Éditions Matériologiques. Mis en ligne le 7 octobre 2010. URL : www.materiologiques.com

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.

Michel Paty **Le matérialisme considéré du point de vue cosmologique et évolutionniste¹**

J'ai choisi pour titre de cet article «Le matérialisme considéré du point de vue cosmologique et évolutionniste». Mais j'aurais pu choisir aussi bien la formule inverse, qui aurait tout autant fait du sens : «La cosmologie et l'évolution considérées d'un point de vue matérialiste»... En fait, le point de vue «cosmologique et évolutionniste» scientifique, caractéristique de la science d'aujourd'hui, est sous-tendu par une perspective matérialiste (celle que l'on peut appeler «matérialisme scientifique», suivant laquelle la matière, c'est-à-dire ce qui constitue le monde, est connaissable par la science, sans restriction ou impossibilité *a priori*). Or, en même temps, ce point de vue scientifique, de la science actuelle, oblige le matérialisme à ne pas être réducteur, et à faire toute leur place aux spécificités des divers états d'organisation (des divers niveaux) de la matière inanimée, des formes vivantes et des êtres pensants. Cette double (et dialectique) considération forme l'orientation de mon propos.

Un premier thème d'intérêt, dans cette perspective, est celui de la *constitution «naturelle» de la connaissance* sur le monde par les êtres pensants qui en sont issus et qui continuent de lui appartenir, qui

1. Texte original de la conférence en espagnol à la *Jornada de estudios materialistas: «Introducción a la Historia del Materialismo»*, Archivo y Biblioteca-Chile en Suecia, Librería Latino-Americana, Stockholm (Suède), 15 juin 2007.

s'accompagne de la réflexion sur lui, réflexion que, dans notre culture, nous appelons «philosophique». Nous posons aujourd'hui les problèmes suscités par la mise en rapport de ces connaissances et de cette réflexion en termes de «*science et philosophie*». Aux diverses étapes de l'évolution et de l'histoire des êtres humains et des sociétés, ces rapports ont pu se poser de diverses manières. L'humanité n'a pas attendu qu'advienne «la science», au sens où nous l'entendons aujourd'hui, pour s'adonner à la connaissance des objets et des phénomènes du monde, pour en former de nouveaux qui portent sa marque, et pour concevoir des modes de relation entre les humains et le reste du monde. Des savoirs «objectifs» sur la nature ont été obtenus, insérés dans des systèmes de croyance, formés à partir des dimensions psychologique et sociale (sur une base biologique) des êtres humains vivant dans des groupes ou des peuplades. L'immanence de notre condition, qui caractérise le point de vue matérialiste, nous rend sensibles à cet aspect : la connaissance pratique, théorique, de la matière du monde, et la réflexion sur elle et sur le monde en général, et sur la situation de l'homme dans le monde, accompagne l'hominisation et la socialisation, en les chargeant progressivement de contenu, par les acquis successifs et les transformations (qui comprennent aussi des ruptures entre des systèmes de pensée). Ceci, de la préhistoire à la science contemporaine (qui se veut résolument «objective»), en passant par des systèmes très divers de connaissance (mythiques, etc.).

Il faudrait maintenant indiquer *ce que nous apprend la science contemporaine*, et ce ne sera fait ici qu'en très peu de mots, suffisamment évocateurs : la structure de la matière, l'organisation du vivant, la matérialité de la pensée (ses bases matérielles neuronales), les particularités de l'intersubjectif et du social qui font l'histoire humaine (et qui contribuent aussi à la constitution de la pensée). La connaissance scientifique vise à saisir l'*objectivité* du monde, c'est-à-dire la matérialité du monde sous toutes ses formes, l'unité de la matière sous la diversité de ces formes, mais aussi son caractère évolutif. Ce caractère, on le connaissait déjà pour les sociétés, par l'histoire, ainsi que pour les êtres vivants, par leur genèse et leur destinée (concernant les individus, de la naissance à la mort). On le sait aussi depuis assez peu de temps pour la «biosphère» (évolution des espèces) et pour l'Univers dans sa totalité (cosmologie évolutionniste), sans oublier le lieu qui nous a donné naissance et nous héberge, notre niche, notre monde (naguère

«le Monde» signifiait la Terre, des Anciens à Magellan et à Jules Verne) ou notre vaisseau (le «vaisseau Terre») dans l'océan cosmique (la Terre se transforme, comme d'ailleurs tous les objets célestes : voir la science récente de la tectonique des plaques océaniques et continentales du manteau terrestre).

Dans cet ensemble, dans ce tableau de notre connaissance, il y a un *agencement naturel entre tous les états d'organisation et de formes de la matière*, de la matière inanimée à l'homme pensant et social, même si la venue de ce dernier apparaît assez «improbable» (pour cause du caractère unique, singulier – pour l'instant et concernant une connaissance tout empirique – de l'apparition de la vie) : le fait est que cet événement a eu lieu, en un endroit au moins de l'Univers, et que l'on est fondé à lui assigner des causes naturelles, c'est-à-dire matérielles, qui font actuellement l'objet d'actives recherches à la jonction de plusieurs disciplines scientifiques (chimie, biologie, planétologie, «exobiologie», etc.). Cet événement est en droit possible ailleurs, sans doute suivant des modalités variées, mais qui doivent bien se rattacher, d'une manière ou d'une autre (en raison de l'unité de la matière), à celle que nous avons constatée pour nous-mêmes.

Et la pensée aussi est produite (et, avec elle, le social) dans et par la matière qui fait le monde. D'où, entre parenthèses, la légitimité et la nécessité de sciences pour tous les étages ou niveaux de l'organisation de la matière, inanimée, vivante, pensante.

Le fil directeur à travers tout ceci, c'est bien l'unité de la matière du monde, d'où celle (de principe) de la connaissance de la matière et du monde, c'est-à-dire l'unité de la science. Mais cette dernière n'est pas acquise, ni fermée : elle ne s'obtient pas par réduction des sciences les unes aux autres (toutes sont des constructions de la pensée humaine, de nature symbolique, et aucune ne peut prétendre à un point de vue définitif et absolu). L'unité, sinon de la science, du moins entre les sciences, se laisse deviner plutôt, par la cohérence des emboîtements des représentations des différents niveaux d'organisation de la matière. Il semble, au fur et à mesure que les connaissances scientifiques avancent, malgré la grande diversité des objets et des approches, que cette cohérence soit de plus en plus étroite, et c'est bien ce que l'on désire d'un point de vue matérialiste. Autrement dit, on fait de l'unité de la science un programme : on postule que la science (pourvu qu'elle dispose des conditions de son développement, qui comprend la liberté de pensée et

une volonté commune au niveau social) est toujours en mouvement de progrès vers son unité, localement et globalement.

L'humanité ainsi se forme progressivement une connaissance objective du monde, en termes d'objets, de phénomènes, de niveaux d'organisation ou de structuration, par le moyen de concepts, et selon des conditions et conditionnements qui sont de nature matérielle, technique, culturelle, socio-politique, etc. Cette formation de connaissances s'effectue sur le mode du *symbolique*. Par nature, il n'y a pas d'identification «ontologique» entre la connaissance et le monde (Spinoza disait : «Le concept de chien n'aboie pas.») Dans le travail de connaissance, il est donc nécessaire d'éviter les réductionnismes (d'objets de sciences à d'autres), qui correspondraient, en fait, à une identification parfaitement abusive et illusoire entre la matière et notre description de celle-ci. Des catégories de pensée (philosophiques) sous-tendent cet effort de l'intelligence en vue de la connaissance par les concepts et les systèmes structurés de concepts que sont les théories : elles constituent autant d'exigences qu'il importe de tenir ensemble, et éventuellement de redéfinir à cet effet. Parmi elles : unité, spécificités, variétés, émergences, adéquation, objectivité, vérité, réalité... et, bien entendu, matière.

Appartenance au monde et recul ou distanciation. On a souvent opposé, dans l'histoire de la philosophie, la matière à l'esprit, l'immanence à la transcendance, la nécessité à la liberté, l'objectivité aux valeurs. Et où trouver place, dans une philosophie matérialiste, pour la notion de signification ? *Matière, liberté, signification, valeurs* : questions traditionnelles, ou qui surviennent immédiatement à partir de ce que nous avons dit précédemment. La matière nie-t-elle l'esprit ? la réponse est non, bien sûr ! Il faudrait développer sur de telles questions un travail de raisonnement philosophique approfondi. Je m'en tiendrai ici à quelques remarques, proposées de manière synthétique comme des sortes d'aphorismes.

Prenons la «transcendance» : envisagée d'un point de vue matérialiste (au lieu de la nier), elle ne serait plus conçue comme en dehors du monde, mais comme *prise de recul* par rapport à l'immersion dans le monde, cette prise de recul étant le propre de l'homme. L'homme, être vivant et pensant, produit par le monde matériel, mais avec cette propriété de pouvoir se mettre à distance, et qui, par là, découvre et instaure sa *liberté*.

Que serait, par contraste, l'immersion dans le monde sans recul ? Certaines organisations sociales et systèmes de croyances nous en fournissent des exemples. La pensée des sociétés chamaniques ou totémiques ne conçoit pas de séparation entre le sujet et l'objet, entre ce que nous appelons la raison et les affects. La plupart des sociétés anciennes, sinon toutes, ne dissociaient pas la raison et le mythe. Cela vaut aussi pour les sociétés « historiques » disposant de l'écriture : dans la Babylone du dieu Baal, dans la Grèce ou la Rome antiques avec leurs dieux de l'Olympe, ou dans les sociétés des anciens Mexicains qui pratiquaient les sacrifices humains, la destinée des individus (et des sociétés) était marquée par une fatalité sans issue, caractéristique de cette immersion totale. C'est aussi, sous d'autres formes, le cas de sociétés anciennes – mais aussi de certaines contemporaines – régies par le seul pouvoir, contraignant, de la force guerrière, et, plus près de nous, de l'État totalitaire (qui règne par la force policière et militaire aidée du consensus, c'est-à-dire de la soumission admise). Et encore, sous les oripeaux d'un individualisme de façade, dans nos sociétés en cours de « mondialisation », le pouvoir et le primat de l'argent, et l'économie libérale de marché érigée comme référence (le dieu Marché), comme contrainte incontournable des sociétés humaines : mais ce n'est là, nous le savons, qu'une idéologie revêtue des apparences d'une prétendue science.

Sur ce point, nous concluons donc : *transcendance* et *liberté* ne sont pas en opposition avec le *point de vue matérialiste*, entendues dans le sens que nous avons esquissé, et il semble même que ce soit tout le contraire. D'ailleurs, on constate également qu'elles font partie des réalités de notre monde : on les voit se manifester en acte, par-delà les témoignages individuels, dans la société et dans l'histoire des peuples (voir, par exemple, ce que représente la figure d'Antigone, pour ce qui est de l'Antiquité ; ou, d'une manière générale, les héroïsmes, les refus, tout au long de l'histoire).

En même temps que l'expérience de la liberté, l'homme propose la *signification* et énonce notamment que *connaître a un sens*. La connaissance implique l'idée de sens, même si les deux ne se confondent pas. *Comprendre*, c'est, au-delà de *connaître*, au-delà des savoirs acquis, *former du sens*, et c'est lorsque nous avons intériorisé ce sens que les contenus de connaissance nous deviennent *intelligibles* : nous nous les sommes *appropriés* dans nos pensées. Il en résulte une diversité des manières de comprendre, selon les particularités individuelles des

personnes, autour d'un même noyau d'objectivité. De là résultent des exigences à prendre en compte dans l'enseignement. De là vient aussi la possibilité d'invention et de création, en science comme en art.

On peut aussi envisager le *sens* dans une acception plus globale que les connaissances ou compréhensions particulières, locales. Ce sens-là n'est pas élucidable en totalité et sa formulation paraît excéder ce que l'on peut en dire ou analyser. Il est « métaphysique », parfois illusoire et parfois non. (La conscience d'être est indicible : naissance et justification de la poésie et de l'art.)

Précisément, *connaître* et *comprendre* montrent qu'il est possible de former l'idée d'un sens, même dans cette dimension totalisante et valorisante qui touche à la métaphysique, sans le rapporter à un Dieu. Pour Camus, avec ou sans Dieu, le monde est absurde : qu'un tel jugement soit possible fait assez voir que le sens ne s'identifie pas à un Dieu. Le sens ou l'absurde sont des jugements qui se situent en dehors de la matière nue, certes, mais nous pouvons les formuler sans le secours d'aucune Révélation, par nos propres moyens (ceux de l'intelligence humaine).

Ces jugements manifestent tout simplement le recul que l'homme prend par rapport à la matière qui le constitue. La *pensée* a ce pouvoir de réflexion sur elle-même (de *réflexivité*), dont une forme intense est la *conscience*, qui est la référence de l'*éthique* et de la *responsabilité*, et qui est sous-jacente au *droit*, régulateur des sociétés, du moins celles qui font à un degré ou un autre référence à la conscience et à la responsabilité. D'une manière générale, ce que l'on appelle le « spirituel » peut être approché de cette manière immanentiste, sans être rabaissé, et en restant valorisé.

On peut chercher des raisons naturelles à l'existence de cette dimension de *transcendance dans l'immanence*, à partir de la seule matière (bien au-delà du simple arrangement des neurones : par exemple, la révolte de l'individu ou le chagrin devant la mort, qui lui en fait prendre le contre-pied, jusqu'à inventer des dieux : rites et croyances dont la base est avant tout sociale). Mais il reste que ces croyances, et les actions qui s'y rapportent, sont formées, formulées, communiquées, reprises et choisies par l'être humain matériel, et constituent désormais une référence pour lui, référence de signification, d'action ou de justification (d'où la formation historique des *valeurs*).

D'autres questions mériteraient d'être évoquées et discutées dans la même perspective. Je m'en tiendrai, pour conclure, à deux remarques sur *ce qu'est* ou *n'est pas*, en fin de compte, de manière fondamentale, le matérialisme, comme position philosophique.

Le matérialisme n'est pas, et ne peut être une religion, et celui qui professe une philosophie matérialiste ne peut se comporter comme le détenteur d'une foi doctrinaire et présenter sa philosophie comme un dogme. (Foi ou dogme, elle ne serait plus une philosophie, qui suppose raisonnement et liberté de l'esprit.) C'est un *choix* volontaire, raisonné, portant sur une approche de la signification du monde et de notre situation dans le monde. Ce choix ou cette décision suppose une *pensée critique*, qui implique *l'idée de sujet*, la *conscience d'être sujet* (réflexivité, et transcendance : conscience) et une idée de la *liberté* : sujet et liberté sont soutenus par leur matérialité, qui les pose dans le monde. Les catégories qui sous-tendent notre réflexion ne sont pas figées : nous devons les aménager pour que ceci soit possible. Sans cette remise en question, notre pensée du monde et de nous-mêmes pourrait être contradictoire ou même monstrueuse.

Le matérialisme n'est pas une science : c'est une position avant tout philosophique, et aussi métaphysique, ouverte sur les apports des sciences, sur *la science* dans sa diversité et dans son unité, et qui permet de concevoir le monde dans son immanence.

🔗 Notice biographique 🔗

Michel Paty est directeur de recherche émérite au CNRS, Laboratoire SPHERE du CNRS et de l'Université Diderot-Paris 7 (UMR 7219). Site avec textes disponibles à télécharger.

Avertissement. Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur et des auteurs. Les textes et illustrations figurant sur ce site peuvent être consultés et reproduits sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage personnel, scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document selon ce modèle :

Michel Paty, «Le matérialisme considéré du point de vue cosmologique et évolutionniste», in Gérard Lambert et Marc Silberstein (dir.), *Matière première. Revue d'épistémologie [en ligne]. Nouvelle série, N° 1/2010 : Épistémologie de la médecine et de la santé*, Éditions Matériologiques. Mis en ligne le 12 octobre 2010. URL : www.materiologiques.com

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.

