L'éthique évolutionniste : de l'altruisme biologique à la morale

Clavien, Christine

L'éthique évolutionniste : de l'altruisme biologique à la morale

Thèse réalisée en cotutelle aux

Institut de philosophie
Faculté des lettres et sciences humaines
Université de Neuchâtel
&
Institut d'Histoire et de Philosophie des Sciences et des Techniques
Université de Paris I Panthéon-Sorbonne

Présentée pour l'obtention du grade de docteur ès Lettres

par

Christine Clavien

Acceptée sur proposition du jury :

Prof. **Daniel Schulthess**, directeur de thèse
Prof. **Jean Gayon**, directeur de thèse
Prof **Daniel Andler**, rapporteur
Prof **Ronald de Sousa**, rapporteur
Prof **Nicolas Perrin**, rapporteur
Prof **Dan Sperber**, rapporteur

Soutenue le 11 janvier 2008
Université de Neuchâtel
2008

Université de Neuchâtel

Faculté des lettres et sciences humaines

La doyenne
- Espace Louis-Agassiz 1
- CH-2000 Neuchâtel

IMPRIMATUR

La Faculté des lettres et sciences humaines de l'Université de Neuchâtel, sur les rapports de M. Daniel Schulthess, co-directeur de thèse, professeur ordinaire de philosophie à l'Université de Neuchâtel ; M. Jean Gayon, co-directeur de thèse, professeur à l'Université de Paris I-Panthéon Sorbonne ; M. Daniel Andler, professeur à l'Université de Paris-Sorbonne, Paris IV ; M. Ronald de Sousa, professeur émérite de l'Université de Toronto ; M. Nicolas Perrin, professeur à l'Université de Lausanne ; M. Dan Sperber, directeur de recherches au CNRS, Paris, autorise l'impression de la thèse présentée par Mme Christine Clavien, en laissant à l'auteur la responsabilité des opinions énoncées.

Neuchâtel, le 11 janvier 2008

La doyenne
Ellen Hertz

■ Téléphone : +41 32 718 17 00 ■ Fax : +41 32 718 17 01 ■ E-mail : doyen.lettres@unine.ch ■ www.unine.ch/lettres

Mots clés en français :

altruisme, coopération, culture, émotion, éthique, éthique évolutionniste, évolution, jugement de valeur, métaéthique, morale, motivation, norme, réalisme moral, réciprocité, sélection de groupe, sélection de parentèle, sentimentalisme, sophisme naturaliste, théorie de l'erreur, théorie de l'évolution, théorie des jeux, valeur

Mots clés en anglais :

altruism, cooperation, culture, emotion, ethics, evolutionary ethics, evolution, value judgment, metaethics, morality, motivation, norm, moral realism, reciprocity, group selection, kin selection, sentimentalism, naturalistic fallacy, error theory, theory of evolution, game theory, value

Résumé :

Cet ouvrage est une contribution à l'éthique évolutionniste, un courant de pensée dont les débuts remontent à la seconde moitié du 19è siècle et qui s'inspire des connaissances de la biologie de l'évolution pour aborder les questions éthiques (genèse de la moralité, métaéthique, fondement des principes moraux). L'éthique évolutionniste comprise dans son sens contemporain est un domaine de pensée à la fois riche et complexe, imprégné non seulement des idées de Darwin mais également de données et outils théoriques issus de sciences aussi variées que la biologie de l'évolution, la théorie des jeux, la psychologie, la neurologie, l'anthropologie, l'économie empirique ou les sciences cognitives en général. Ainsi ce courant peut être considéré comme un emblème de l'interdisciplinarité.
Cet ouvrage présente une large palette de théories et données scientifiques sur lesquelles reposent les réflexions menées en éthique évolutionniste. Il montre également comment ce courant s'inscrit à l'intérieur même de la philosophie morale sans se résumer à une seule ligne de pensée. Il s'agit plutôt d'une méthodologie originale dont il est indispensable de clarifier l'application. C'est d'ailleurs l'objectif essentiel de l'ouvrage : mesurer les limites et les possibilités de l'adoption d'une perspective évolutionnaire et scientifique dans le domaine moral. Il apparaît que dans le foisonnement de voies possibles, seules certaines résistent à la critique.

English title : Evolutionary Ethics: from biological altruism to morality

Summary :

This work is a contribution to evolutionary ethics, a particular approach to ethical questions (origins of morality, metaethics, foundation of moral principles) that goes back to the second half of the 19th century and is inspired by evolutionary biology. Nowadays, evolutionary ethics has become a rich and complex interdisciplinary field that rests not only on Darwin's ideas but on a vast range of recent developments in research fields such as evolutionary biology, game theory, psychology, neurology, anthropology, empirical economics, and cognitive sciences in general.
This volume covers the main scientific theories and empirical data relevant to evolutionary ethics. It also shows how this field remains within the scope of moral philosophy without coming down to one single line of thought; it is best understood as a methodology, a new way of grasping the phenomenon of morality. The main aim of this writing is to define the proper use of this methodology, that is, to identify the limits and possibilities of an evolutionary and scientific approach to morality. It appears that among the wide variety of possible views, few resist criticism.

Remerciements

Mes premiers remerciements vont à Daniel SCHULTHESS et Jean GAYON, mes directeurs de thèse dans le cadre d'une cotutelle entre les Universités de Neuchâtel et de Paris I. Leur soutien et leurs encouragements ont été essentiels durant les années de préparation de cette thèse. Je leur doit également à tous deux d'avoir orienté mes recherches, de m'avoir conseillée dans mes lectures, et enfin d'avoir lu et commenté dans le plus grand détail les différentes versions préliminaires de cet écrit. C'était un réel défi de satisfaire à la demande de précision dans les explications imposée par Jean GAYON et à la rigueur argumentative exigée par Daniel SCHULTHESS. J'espère avoir répondu au mieux à leurs attentes.

Les idées présentées dans ce travail ont germé dans les murs de différentes Universités. Il y a d'abord l'Université de Neuchâtel où j'ai été engagée comme assistante doctorante. Il y a également l'Université de Lausanne où j'ai eu l'occasion de mener des projets interdisciplinaires dans le cadre du groupe « déterminismes et libertés ». J'ai en outre bénéficié d'excellents contacts avec les biologistes de Lausanne, en particulier Nicolas PERRIN qui a lu et commenté différentes versions préliminaires de la partie biologique de ce travail. Mais également avec Laurent KELLER, toujours prêt à se lancer dans de longs échanges d'idées. C'est à eux que je dois d'avoir compris le mécanisme de la sélection de parentèle et les limites de la sélection de groupe.

De fréquentes visites tout au long de ma thèse ainsi qu'un séjour d'un semestre à l'IHPST (Institut d'Histoire et de Philosophie des Sciences et des Techniques) de Paris m'ont fourni une excellente initiation à la philosophie de la biologie. Je suis reconnaissante à toute l'équipe de l'IHPST pour leur accueil chaleureux et en particulier à Philippe HUNEMAN qui c'est admirablement prêté à contrer mes idées.

Une gratitude toute particulière va à Peter GOLDIE qui m'a réservé le meilleur accueil à l'Université de Manchester. C'est avec les larmes aux yeux que je suis revenue en Suisse au terme d'un semestre intense durant lequel cet homme d'exception m'a initiée à la philosophie des émotions et fait bénéficier de ses critiques et conseils avisés. Merci également à la joyeuse équipe de doctorants de Manchester avec laquelle j'ai pu échanger tant d'idées et de bons moments.

Je suis également extrêmement redevable à mes jurés de thèse, Daniel ANDLER, Ronald DE SOUSA, Nicolas PERRIN et Dan SPERBER pour leurs critiques constructives à la fois dans leurs rapports et au cours de la défense.

Mes recherches et séjours à l'étranger ont été possibles grâce aux fonds qui m'ont été octroyés par l'Université de Neuchâtel (assistanat), l'Université de Lausanne (via Anthropos) et la CRUS (bourse de cotutelle).

En outre, j'aimerais remercier les personnes dont les précieux commentaires ont contribué substantiellement à mes recherches : Nicolas BAUMARD, Vasco CASTELA, Michel CHAPUISAT, Eric CHARMETANT, Fabrice CLEMENT, Julien DEONNA, Ronnie DE SOUSA, Jacques DUBOCHET, Michael ESFELD, Luc FAUCHER, Chloe FITZGERALD, Laurence KAUFMANN, Christian MAURER, Jérôme RAVAT, Fabrice TERONI, Marco TOMASSINI.

Merci également aux étudiants de l'Université de Neuchâtel sur lesquels j'ai pu tester mes idées dans le cadre de deux séminaires avancés, l'un sur l'altruisme et l'autre sur la métaéthique évolutionniste.

Enfin je suis redevable à mes premiers lecteurs pour avoir traqué les fautes et coquilles de cet écrit : Michel CLAVIEN, Jacques MONNIER, Corinne PERRITAZ, Patrick PERRITAZ.

Table des matières

Introduction

Ces dernières décennies ont vu des changements considérables dans les sciences du vivant. Il est difficile d'imaginer que ces avancées scientifiques soient sans influence sur la réflexion éthique. C'est du moins l'avis de l'éthique évolutionniste,[1] un courant de pensée qui s'oppose à une marginalisation du domaine des faits par rapport au domaine moral. Les défenseurs de ce courant se gardent d'accorder une trop grande autonomie à l'être humain par rapport à l'ordre naturel ;[2] au contraire, ils s'efforcent de montrer que le comportement humain dépend en bonne partie de la nature.

A première vue, ce projet d'insertion de l'homme dans la nature fait de l'éthique évolutionniste un courant philosophique troublant. Mais en réalité l'idée n'est pas nouvelle ; dans l'antiquité déjà, il était courant d'aborder les questions d'éthique en rapport avec les explications sur l'origine de l'univers et de l'homme. Simplement, le type d'approche développé par l'éthique évolutionniste connaît une histoire plus courte. Ce n'est pas des écrits d'ARISTOTE ou de PLATON qu'elle s'inspire mais des travaux de Charles DARWIN.

En 1859, Charles DARWIN publie *L'origine des espèces*, dans lequel il présente sa théorie de l'évolution des espèces et soutient le point de vue que l'*homo sapiens*, tout comme les autres espèces animales, n'est autre qu'un produit naturel de l'évolution. Environ dix ans plus tard, DARWIN achève la rédaction de *La filiation de l'homme*, ouvrage dans lequel il tire les conclusions de sa théorie de l'évolution dans le domaine de la morale. Pour lui, ce qui distingue l'homme de l'animal est une affaire de degré et non de nature. Il est vrai qu'il considère l'homme comme la créature naturelle la plus évoluée, capable de raisonner, communiquer et penser en termes de ce qui est bien pour la communauté ; ces facultés hautement développées lui permettent d'étendre sa

[1] Afin de suivre une terminologie plus largement utilisée, dans cet ouvrage, je ferai usage de la terminaison 'iste' pour les théories et sciences inspirées d'une approche évolutionnaire. Ainsi, j'utiliserai « éthique/psychologie/etc. évolutionn*iste* » plutôt que « éthique/etc. évolutionn*aire* ».

[2] Pour ne mentionner qu'un exemple, dans les débats sur le positionnement de l'éthique par rapport à la science, Samuel WILBERFORCE est un avocat de la thèse de l'autonomie humaine par rapport à l'ordre naturel. Il est connu pour avoir contesté avec véhémence les idées de DARWIN. Voici l'extrait d'un de ses sermons : « La suprématie dont l'homme a hérité sur la terre ; le pouvoir humain du discours articulé ; le don humain de la raison ; le libre arbitre et la responsabilité humaine (…) – toutes ces choses sont absolument inconciliables avec la notion dégradante de l'origine de brute que l'on attribue à celui qui fut créé à l'image de Dieu. » (WILBERFORCE 1874, pp. 94-95, cité dans DENNETT 2000/1995)

sympathie naturelle au-delà du cercle restreint de sa tribu et de développer un véritable « sens moral ». Mais cela n'exclut pas que les primates deviennent un jour des êtres moraux. Au fond, la moralité est un simple produit de l'évolution. Elle ne porte aucun signe de noblesse particulier, si ce n'est peut-être que seuls des individus aux capacités cognitives hautement développées peuvent la pratiquer.[3]

Du vivant de DARWIN déjà, ces idées ont été reprises, réinterprétées et développées dans différentes théories morales, si bien que l'on peut parler de la naissance d'un véritable courant de pensée : l'éthique évolutionniste. Ce nouveau courant a connu une histoire triplement malheureuse. A ses débuts, sous le nom de « darwinisme social », il a occasionnellement servi à défendre la cause du laisser-faire en matière de politique sociale. L'instigateur de ce genre de théories est Herbert SPENCER, un contemporain de DARWIN ; en s'appuyant sur le principe de la survie du plus apte, il prônait l'élimination ou l'abandon des plus faibles au nom de la survie des plus aptes. L'éthique évolutionniste a ensuite été mise au service de la cause eugéniste[4] dans la première moitié du 20è siècle (notamment aux Etats-Unis et en Allemagne). Enfin elle a suscité une incroyable polémique dans les années 70 lorsque le sociobiologiste Edward O. WILSON s'est pris à vouloir retirer la morale des mains des philosophes. Dans *Sociobiology: The New Synthesis*,[5] Edward O. WILSON suggère que

> « les scientifiques et les humanistes devraient envisager la possibilité que le temps est venu de retirer temporairement l'éthique des mains des philosophes et de la biologiser » (O. WILSON 1975, p. 562, ma traduction).[6]

Cette formule pour le moins polémique a provoqué un tollé général dans le monde des sciences humaines (voir notamment SAHLINS 1980/1976 ; THOMPSON 1999 ;

[3] Pour un bref exposé de la théorie morale de DARWIN, voir RICHARDS 1999.

[4] Francis GALTON, le cousin de DARWIN, a élaboré ce terme pour désigner la science des conditions favorables à la reproduction humaine.

[5] Avec cet ouvrage, Edward WILSON pose les bases d'un nouveau courant en biologie : la sociobiologie, dont le projet est d'expliquer les comportements et les règles sociales directement en termes de sélection naturelle et génétique.

[6] « Scientists and humanists should consider together the possibility that the time has come for ethics to be removed temporarily from the hands of the philosophers and biologicized. » (E. WILSON 1975, p. 562)

FARBER 1994).[7] De fait, le conseil d'Edward WILSON n'a guère été suivi puisque peu de biologistes se sont pris d'affection pour les questions traditionnellement traitées en philosophie morale. En revanche, on lui doit probablement d'avoir causé un regain d'intérêt pour l'éthique évolutionniste ; un bon nombre de philosophes se sont ouverts au projet d'une réelle naturalisation de la morale.

Aujourd'hui, les adeptes de l'éthique évolutionniste ne considèrent pas leur approche comme une option distincte des philosophies morales traditionnelles ; ils cherchent simplement à y introduire le point de vue évolutionnaire. En outre, ils ne se contentent pas de trouver leur inspiration dans les seules idées de DARWIN ; ils utilisent les données et les outils théoriques de tout un ensemble de sciences contemporaines telles que la biologie de l'évolution, la théorie des jeux, la psychologie, la neurologie ou l'anthropologie. Ainsi l'éthique évolutionniste, qui ne porte peut-être plus si bien son nom étant donné le large spectre de sciences dont elle s'inspire, est devenue un petit paradis de l'interdisciplinarité où les chercheurs de tous bords mettent en commun leurs idées et leurs connaissances, si bien qu'il est fort difficile de se faire une idée globale des différentes positions défendues dans le cadre de ce courant de pensée.

Dans cet ouvrage, je ne m'attarderai pas davantage sur l'histoire de l'éthique évolutionniste, pas plus que sur l'analyse détaillée de théories particulières défendues par tel ou tel penseur contemporain. Mon propos sera plus général ; je me demanderai dans quelle mesure il peut s'avérer fructueux de prendre le point de vue de l'évolution et de tenir compte d'un large spectre de données empiriques pour aborder les questions d'éthique. En d'autres termes, il s'agira de sonder les limites et les possibilités d'une éthique évolutionniste.

Dans le cadre de mes recherches, deux livres ont tout particulièrement marqué ma pensée et orienté mes recherches ; de ce fait, ils ont largement influencé le contenu et la construction même de cet ouvrage. Le premier, *Unto Others,* est l'œuvre d'Elliott SOBER et David WILSON. Ils y présentent une distinction extrêmement utile entre l'*altruisme évolutionnaire* et l'*altruisme psychologique,* proposant d'analyser séparément ces deux phénomènes logiquement distincts. Dans la première partie de cet écrit, je reprends cette terminologie et, m'inspirant de la structure de leur ouvrage, je

[7] Notons que si l'on peut reprocher à E. WILSON son penchant pour la polémique, il n'en demeure pas moins un biologiste de l'évolution hors pair.

consacre un chapitre à chacune de ces deux formes d'altruisme. En revanche, je m'éloigne sensiblement de leur position dans le détail de l'analyse. Mon second livre de référence a été *Wise Choices, Apt Feelings*, écrit par Allan GIBBARD. Même si l'auteur n'utilise pas ce terme, il s'agit d'un magnifique projet d'éthique évolutionniste qui s'articule à la fois aux niveaux descriptif, métaéthique et normatif. Dans la deuxième partie de cet écrit, j'applique la méthodologie de GIBBARD qui consiste à partir d'une analyse descriptive fouillée de l'activité et de la pensée morale pour en tirer des conséquences aux niveaux métaéthique et normatif. Son influence n'apparaîtra cependant qu'en filigrane puisque je ne discute que brièvement sa position métaéthique pour la remettre en question.

Cet ouvrage comporte deux parties : l'une est essentiellement centrée sur les nouveaux développements des différentes sciences évolutionnaires, l'autre entre dans l'élaboration d'une éthique évolutionniste à proprement parler.

La première partie traite essentiellement de la question de l'évolution de l'altruisme. Cette orientation thématique est due au fait que traditionnellement, les éthiciens évolutionnistes cherchent à expliquer la morale en termes d'altruisme ; ils pensent que la genèse de la moralité est à trouver dans celle de l'altruisme. Or si l'on veut développer une théorie de ce type qui soit crédible, il faut montrer qu'elle repose sur des recherches sérieuses et non sur des hypothèses vagues et hâtives. C'est la raison pour laquelle je consacre de longues pages à l'explication des comportements coopératifs et altruistes chez les animaux et les êtres humains. Au fond, j'aimerais éviter les critiques que l'on peut adresser à ce que Frans DE WAAL appelle la « sociobiologie de vulgarisation »

> « La sociobiologie de vulgarisation pose un problème d'ordre général : elle traite les questions complexes d'une façon tellement condensée que, même lorsque les auteurs sont parfaitement conscients de ce qu'ils laissent de côté, les lecteurs n'ont aucun moyen de le savoir. Ces simplifications sont ensuite reprises à satiété par les auteurs moins bien informés jusqu'à ce qu'elles se répandent dans toute la discipline, de sorte qu'on est alors obligé de les attaquer comme s'il s'agissait d'idées sérieuses. » (DE WAAL 1997/1996, p. 276)

Plusieurs générations de biologistes de l'évolution ont consacré leurs recherches à la question de l'évolution de l'altruisme et, depuis une vingtaine d'années, les théoriciens des jeux, économistes, psychologues et anthropologues d'obédience évolutionnaire se sont intéressés à la même question. La première partie de cet ouvrage retrace les controverses qui ont fait rage dans ce domaine de recherche. Je tâcherai de mettre en évidence les hypothèses les plus plausibles et de dégager les mécanismes sous-jacents à l'évolution des comportements altruistes et plus généralement coopératifs.

L'ensemble de cette première partie s'articule autour d'une double distinction fondamentale. La première distinction se fait entre l'« altruisme évolutionnaire » et l'« altruisme psychologique ». Si l'on veut aborder les questions liées à la notion d'altruisme à la fois par le biais de la philosophie et par celui des sciences évolutionnaires, on se trouve rapidement confronté à des problèmes de définition. En effet, il y a deux manières de comprendre la notion d'altruisme ; la version *évolutionnaire* se distingue nettement de la conception ordinaire que nous nous en faisons. Les théoriciens de l'évolution définissent l'altruisme en termes de valeur de survie et de reproduction (*fitness*) : un comportement est altruiste s'il a pour *effet* d'augmenter la *fitness* d'autrui aux dépens de sa propre *fitness*. Ce sont donc les conséquences des comportements dont on tient compte pour déceler l'altruisme. En revanche, dans le domaine du sens commun tout comme en psychologie et en philosophie, on considère généralement qu'un acte est altruiste s'il a été causé par une *motivation* dirigée vers le bien d'autrui. En ce sens, on parle d'« altruisme psychologique ». Cette distinction est importante, car si l'on peut expliquer la genèse de l'altruisme *évolutionnaire*, on est encore bien loin de celle de la moralité. En effet, même si l'on peut admettre que l'altruisme et la moralité entretiennent des liens étroits (cette idée sera défendue aux chapitres 4 et 5), c'est de l'altruisme *psychologique* dont il est question, c'est-à-dire tel qu'il est conçu par les psychologues et les philosophes. Il s'agit donc d'expliquer l'évolution de ce dernier. Le point intéressant est que, étiologiquement parlant, l'altruisme évolutionnaire semble être une condition nécessaire à l'évolution d'une forme d'altruisme psychologique[8] ; c'est du moins en ce sens que j'argumenterai au chapitre 3. En résumé, si l'on veut expliquer la genèse de la moralité

[8] Il s'agit de la forme motivationnelle de l'altruisme psychologique. Au chapitre 3 (section 3.3.3), je la distingue de la forme *sophistiquée* qui est une activité réflexive particulière.

et si l'on admet qu'elle est intrinsèquement liée à la motivation altruiste, il faut expliquer la genèse de l'altruisme psychologique, laquelle dépend de la genèse de l'altruisme évolutionnaire. L'organisation du présent ouvrage reflète cette logique : je commencerai par les conjectures sur l'évolution de l'altruisme évolutionnaire (chap. 2), puis de l'altruisme psychologique (chap. 3) avant de formuler des hypothèses sur l'origine de la morale (chap. 5).

La seconde distinction autour de laquelle s'articule la première partie de cet ouvrage concerne l'altruisme animal *versus* l'altruisme humain. Ce dernier doit être compris comme une version raffinée du premier. Au niveau de l'explication des comportements coopératifs et altruistes *évolutionnaires,* on retrouve chez les êtres humains les mêmes mécanismes que chez les animaux : la sélection de parentèle, la réciprocité directe et probablement aussi le signal coûteux. Mais les êtres humains se distinguent des animaux par leurs capacités cognitives hautement développées (mémoire, imitation, apprentissage individuel, apprentissage culturel) ; celles-ci leur permettent une exploitation plus efficace de certains mécanismes (réciprocité directe) ainsi que la mise en œuvre de nouveaux mécanismes : la réciprocité indirecte et la sélection culturelle de groupe. Saisir le fonctionnement et l'impact de ces différents mécanismes permet de comprendre pourquoi les êtes humains sont à la fois nettement plus sociaux et plus opportunistes (la contradiction n'est qu'apparente !) que les autres espèces animales. Quant à l'analyse de l'altruisme *psychologique* (dans sa version motivationnelle), il apparaîtra que seuls les individus disposant de certaines capacités cognitives (conscience de soi, théorie de l'esprit) en sont capables. C'est le cas des êtres humains et peut-être de certaines autres espèces animales sociales comme les chimpanzés par exemple. En résumé, l'étude du comportement animal fournit les premiers éléments d'explication de l'altruisme humain, mais ce dernier ne peut être pleinement compris qu'au terme d'une analyse qui tient compte des capacités propres aux êtres humains.

L'objectif majeur de la première partie de cet ouvrage est de préparer le terrain à l'élaboration d'une éthique évolutionniste, laquelle fait l'objet de la deuxième partie. Les premiers chapitres nous fournissent une analyse des conditions nécessaires à la stabilisation de comportements hautement sociaux (parenté, contacts répétés, contrôle social, sanction, régulation normative, etc.), ainsi que des différents biais psychologiques inscrits dans notre nature humaine (biais de contenu, biais du

conformisme ou du prestige, etc.) ; sur cette base, dans le chapitre sur l'éthique descriptive (essentiellement les sections 5.2.1 et 5.2.2), je peux montrer comment ces mécanismes et biais influencent notre activité morale. En outre, deux éléments utiles pour la réflexion éthique sont thématisés dans le chapitre sur l'altruisme psychologique. Premièrement la motivation altruiste (qui se résume en fait à la formation d'émotions altruistes) dépend de la stabilisation évolutionnaire de l'altruisme évolutionnaire et plus généralement de systèmes favorisant la coopération (section 3.4). Deuxièmement, cette version *motivationnelle* de l'altruisme psychologique doit être distinguée d'une version *sophistiquée*, laquelle relève des croyances et réflexions des individus, plus précisément de ce qu'ils considèrent comme étant les motifs de leurs actions (section 3.3.3). Le chapitre sur l'éthique descriptive exploite cette distinction. J'y montre que la forme *motivationnelle* de l'altruisme psychologique (émotions altruistes) est essentielle pour nous motiver à l'action morale ; de plus, en tant que réactions émotionnelles face à certains types de situations, les émotions altruistes fournissent les premières impulsions à la réflexion morale proprement dite. Quant à l'altruisme *sophistiqué*, je suggère qu'il s'agit d'une condition nécessaire à la production d'assertions morales ; il faut même lui accorder le statut de critère de moralité par excellence (sections 4.2.2 et 5.4.1).

Dans la seconde partie de cet écrit, je tente de montrer que les conditions imposées par une approche évolutionnaire de l'éthique permettent l'élaboration d'un système moral crédible, solide, et qui ne met pas en cause le sens commun. Plus précisément, je défends l'idée que la moralité peut être *expliquée* à l'aide de données empiriques et d'outils théoriques issus des théories évolutionnistes, et que l'on peut même recourir à ces derniers dans l'entreprise de *justification* de nos assertions morales.

La façon la plus constructive de concevoir l'éthique évolutionniste est de la comprendre comme une nouvelle manière de pratiquer l'éthique ; il s'agit d'utiliser un outil de réflexion supplémentaire qui n'est autre que l'adoption d'une perspective évolutionnaire. Cet angle d'approche nous incite à effectuer certains choix théoriques plutôt que d'autres et fournit à l'occasion des arguments supplémentaires pour prendre position dans un débat (cela apparaît de manière particulièrement claire au chapitre 6). Ainsi l'éthique évolutionniste est tout sauf un projet de remplacement de la philosophie morale. Elle ne peut pas non plus être considérée comme un nouveau courant au même titre que l'utilitarisme ou le déontologisme ; même si ce n'est pas la voie qui sera choisie ici, il est tout à fait possible pour un éthicien évolutionniste de s'inscrire dans un

courant déontologique (RAUSCHER 1997), utilitariste (R. WRIGHT 1994/1995) ou de souscrire à une éthique de la vertu (ARNHART 1998).

Adopter une approche évolutionnaire est plus ou moins utile en fonction des niveaux de réflexion éthique. Au chapitre 4, je distingue quatre niveaux : i) l'éthique descriptive où il est question de la genèse de la morale et de l'explication de la manière dont les gens pensent et agissent moralement ; ii) la métaéthique où il s'agit de définir le statut ontologique de la morale et d'envisager la possibilité d'une connaissance morale ; iii) l'éthique normative qui traite de la justification et du fondement de nos jugements moraux ; et enfin iv) l'éthique appliquée où l'on cherche à résoudre les conflits moraux existants. Dans cet écrit, seuls les trois premiers domaines de réflexion retiennent mon attention ; mon objectif est de montrer que l'éthique évolutionniste est principalement opérante aux niveaux descriptif et métaéthique et que, dans une moindre mesure, il est également utile de recourir à une perspective évolutionnaire en éthique normative.

Les trois derniers chapitres sont l'occasion d'une prise de position à l'intérieur même de l'éthique évolutionniste. Les principales théories et courants défendus dans ce domaine font l'objet d'une analyse critique.

Au niveau descriptif (chap. 5), si on adopte une approche évolutionnaire, on se fera une certaine conception de la morale ; elle doit être le résultat d'un processus naturel d'évolution. Ainsi, si l'on veut en donner une interprétation convaincante, il faudra recourir à des schèmes d'explications propres aux théories évolutionnistes, telles que la sélection naturelle, l'adaptation, la fonction évolutionnaire, etc. Contrairement à une forte tradition en éthique évolutionniste selon laquelle la moralité est une adaptation (elle aurait été sélectionnée parce qu'elle répond à un besoin, apparu au cours de l'évolution humaine), je défends l'idée que la moralité est un produit dérivé, lequel repose sur des capacités adaptatives. Plus précisément, pour reprendre une formule de Luc FAUCHER (2007, p. 117) il semblerait que la morale est une « espèce pratique », c'est-à-dire une manière dont *nous* catégorisons certains états de fait qui nous intéressent. Dès lors, il est permis de définir le domaine de la moralité de manière relativement arbitraire. A cet effet, je propose de faire appel à la version sophistiquée de l'altruisme psychologique : pour qu'il y ait moralité il faut, de la part du sujet, un effort réflexif qui tienne compte des intérêts et du bien-être d'autrui (section 5.4.1).

Indépendamment de la définition précise de la moralité, l'activité évaluative et normative en général peut faire l'objet d'une analyse descriptive. Les données empiriques et les théories évolutionnistes permettent d'esquisser les grandes lignes de la

manière dont les gens pensent et agissent en termes évaluatifs et normatifs. A la section 5.2, je présente un « tableau affectif » qui distingue un aspect *spontané* et un aspect *réflexif* de cette activité : le premier est composé de réactions émotionnelles guidées par des valeurs intuitives, et le second se décline en jugements sophistiqués, normes et valeurs conscientes. Ce tableau nous permet de saisir le lien entre le fait de poser un jugement (ou intégrer une norme) et la motivation à agir conformément à ce jugement (ou à cette norme) : au contraire de ce qu'affirment la plupart des philosophes de la morale, je ne pense pas qu'il s'agit d'un lien de nature causale où le jugement serait source de motivation, mais d'un rapport indirect qui passe par l'influence de nos réactions émotionnelles (lesquelles relèvent du domaine de l'impulsion ou de l'intuition) sur les assertions morales que nous produisons au niveau réflexif. En effet, il y a de bonnes raisons croire que les émotions (et entre autres les émotions altruistes, c'est-à-dire celles qui nous font réagir face au sort d'autrui) jouent un rôle prépondérant dans notre choix de valeurs et de normes.

Il est important de souligner que tout ce qui a été dit jusqu'à maintenant est d'ordre strictement *descriptif*. Pour le moment, aucune position n'a été prise au niveau *normatif*. Statuer sur ce qu'il faut faire ou ne pas faire nécessite la possibilité de justifier nos assertions morales. La manière la plus tentante d'imaginer comment une telle justification peut procéder est de reproduire en morale le modèle de la connaissance empirique : une assertion est considérée comme justifiée si l'on peut démontrer qu'elle correspond à une réalité morale. Cette stratégie repose sur une théorie métaéthique réaliste cognitiviste selon laquelle une réalité morale existe de manière propre et indépendamment des croyances et attitudes individuelles que nous formons à son sujet ; de plus, selon cette théorie, nous avons les moyens cognitifs pour accéder à cette réalité extérieure. Au chapitre 6, j'argumenterai cependant contre une telle conception de la connaissance morale. Je tenterai de montrer (contre un bon nombre d'éthiciens évolutionnistes !) que prendre l'évolution au sérieux implique le rejet de toute la gamme des positions métaéthiques réalistes : pour des raisons inhérentes à l'approche évolutionnaire, il n'est pas possible d'affirmer l'existence de propriétés morales indépendantes des croyances et attitudes des gens. Au contraire, cette approche nous incite à penser que nous projetons les valeurs morales sur les situations que nous observons. En conséquence, il ne peut pas y avoir de connaissance morale au sens analogue à la connaissance empirique.

L'impossibilité d'un fondement ontologique des normes morales pose la question de la possibilité de justifier nos assertions morales. Certains auteurs répondent par la négative : les émotivistes affirment que la moralité n'est qu'une affaire de sentiments personnels à l'image de nos préférences gustatives et les défenseurs de la théorie de l'erreur prétendent que l'objectivité de la morale est une illusion, si bien qu'il est vain de chercher à justifier nos assertions morales. Je me refuse à une vision si pessimiste. Mon dernier chapitre explore les possibilités de justifier nos assertions morales dans le cadre d'une approche évolutionnaire de l'éthique. Différentes tentatives sont exposées et critiquées. Aucune ne s'avère convaincante car il n'est pas acceptable de déduire le bien moral à partir de prémisses empiriques ; dès lors je propose de repenser la notion même de justification morale, c'est-à-dire d'envisager une épistémologie morale sensiblement différente du modèle appliqué dans le domaine scientifique. La morale n'est pas une affaire de découverte d'une quelconque réalité, pas plus qu'elle ne peut être investiguée au seul moyen des outils de la logique. Mais cela n'exclut pas la possibilité d'une sorte d'objectivité. C'est sur le plan psychologique qu'il faut réfléchir : il importe de savoir comment les gens forment leurs valeurs, normes et jugements moraux et quelle importance ils leur accordent. Dans mon exposé du « tableau affectif » (chapitre 5) je défends l'idée que les gens sont généralement convaincus du bien-fondé et de la portée intersubjective des normes qu'ils prônent (objectivité psychologique) ; de plus, leur nature les incite à évaluer les mêmes états de fait de manière similaire et il existe toute une batterie de biais psychologiques qui permettent aux êtres humains d'accorder leurs convictions au fil de leurs interactions et discussions normatives (objectivité de fait). La morale est donc une affaire de convictions personnelles, d'échanges de points de vue et d'influences mutuelles. Sachant cela, il apparaît que le meilleur outil possible pour soutenir ou justifier le choix de certaines valeurs et normes morales plutôt que d'autres est de recourir au sens commun *et* de le renforcer par les meilleures raisons de croire à ce qu'il nous suggère ; à cet effet, il est possible de faire appel aux théories évolutionnistes et aux données scientifiques. Cependant, il est clair que le type de justification dont il est question n'est valable que d'un point de vue individuel et contextuel. Au fond, le sens commun renforcé ne fait rien de plus que de mettre en perspective et raffermir notre adhésion intuitive à certaines valeurs. Mais il a son importance : une conviction qui ne peut pas être renforcée résiste mal à la critique et le philosophe ne saurait s'en accommoder ; l'exigence de cohérence, cette force interne ressentie par tout être rationnel, le pousse à reconsidérer les éléments qui entrent en

contradiction et relancer la réflexion normative jusqu'à ce que toutes ses convictions morales soient renforcées.

Pour terminer, voici un bref résumé du contenu des chapitres qui composent cet ouvrage. Je commence par introduire les notions de base des théories évolutionnistes en tant qu'elles sont appliquées aux niveaux biologique et culturel. Ce premier chapitre fournira les outils nécessaires pour comprendre les détails de la controverse autour de l'altruisme évolutionnaire qui fait l'objet du chapitre 2. Dans le cadre de l'analyse de cette controverse, je procéderai à un examen critique des différentes explications de l'évolution des comportements altruistes et apparemment altruistes. Il apparaîtra que seuls les comportements relevant de la sélection de parentèle ou de la sélection de groupe peuvent prétendre au titre d'altruisme ; de plus, la seconde forme de sélection semble uniquement opérative au niveau culturel. Le chapitre 3 traite de la controverse au sujet de l'existence de l'altruisme psychologique. Je montrerai qu'elle ne peut être résolue à moins de la reformuler sur la base d'une distinction entre l'« altruisme motivationnel » et l'« altruisme sophistiqué ». Ce chapitre se termine par un examen critique des différentes explications de l'évolution de la motivation altruiste. La deuxième partie de l'ouvrage débute (chap. 4) avec quelques définitions et la mise en place d'une structure théorique sur laquelle reposent les chapitres suivants : je commencerai par distinguer quatre niveaux de l'éthique avant de proposer une première approximation de la morale ainsi que de l'apport potentiel d'une approche évolutionnaire en éthique. Cette introduction sera suivie d'un chapitre sur l'éthique descriptive (chap. 5). Il y sera d'abord question de la genèse de la moralité : je soutiendrai que la moralité est un produit dérivé plutôt qu'une adaptation. Cette analyse sera suivie d'un « tableau affectif », une description de la pensée et de l'activité évaluative et normative. Ce tableau n'étant pas suffisant pour délimiter le champ de la moralité, je me demanderai s'il est possible de le faire au moyen d'une analyse des émotions morales. Cette tentative ne portant pas ses fruits, je finirai par proposer deux critères pour individuer la moralité. Le chapitre 6 est dédié aux questions de métaéthique. J'argumenterai longuement contre les positions réalistes, rejetterai brièvement quelques positions antiréalistes (émotivisme, expressivisme, théorie de l'erreur) avant de défendre une forme simple de projectivisme. Le dernier chapitre traitera d'éthique normative. Je rejetterai les tentatives de passage du factuel au normatif par les voies réductionniste et logique. Je montrerai ensuite l'inutilité de la théorie de

l’erreur avant de proposer un modèle de justification de nos assertions morales qui fait appel au « sens commun renforcé ».

Première partie

L'éthique évolutionniste est née de la conjonction de la philosophie morale et de la pensée darwinienne. Dès lors, si l'on veut réellement en comprendre les objectifs, les thèses défendues et les arguments utilisés, il faut commencer par s'initier à la théorie de l'évolution. En effet, c'est en s'informant des problèmes sur lesquels des générations de biologistes ont travaillé que l'on saisira pourquoi les éthiciens évolutionnistes ont focalisé leurs réflexions sur certaines thématiques plutôt que d'autre. C'est en comprenant les détails des systèmes théoriques développés dans le cadre des théories évolutionnistes (dans leurs versions biologique, anthropologique, psychologique et économique) que l'on pénètrera la pertinence et la portée des arguments utilisés en éthique évolutionniste. Pour ces raisons, cet ouvrage fait la part belle aux théories évolutionnistes et en particulier à celles qui sont issues de la biologie darwinienne ; cette dernière pose les bases de tous les arguments et discussions ultérieurs.

Nous verrons au chapitre 2 que la biologie de l'évolution oriente d'emblée la discussion vers l'altruisme. La raison tient à ce que l'existence des comportements altruistes que l'on peut observer dans le monde animal semble remettre en question la théorie de l'évolution elle-même ; cette dernière ne peut être crédible qu'à la condition de pouvoir résoudre le fameux « paradoxe de l'altruisme » (nous verrons qu'un certain nombre de théories permettent de se débarrasser de ce paradoxe). D'autre part, le fait que l'altruisme entretient des liens étroits avec la moralité explique l'intérêt des moralistes pour la biologie de l'évolution. Ainsi, l'altruisme est l'élément clé qui lie l'éthique aux théories évolutionnistes. Une grande partie de cet ouvrage s'occupera donc de la manière dont les biologistes traitent de l'altruisme et du lien entre altruisme et moralité. Ces questions sont complexes et nécessitent un bon nombre de distinctions et de mises en perspective des différentes thèses développées au niveau des théories évolutionnistes d'une part et philosophiques d'autre part.

Une distinction majeure qui sera thématisée tout au long de la première partie de cet ouvrage concerne la notion même d'altruisme qui n'a pas la même signification dans la pensée d'un théoricien évolutionniste ou d'un philosophe ou psychologue. A chacune de ces deux notions d'altruisme sera consacré un chapitre : le chapitre 2 traitera de

l’altruisme *évolutionnaire* alors que le chapitre 3 analysera l’altruisme *psychologique*. Au terme de la première partie de cet ouvrage, les différences et liens entre ces deux notions apparaîtront clairement. Cette discussion nous fournira les moyens théoriques nécessaires à l’analyse de l’impact des théories évolutionnistes dans le domaine de l’éthique, qui fera l’objet de la deuxième partie de l’ouvrage.

1. La théorie de l'évolution et son incidence sur l'auto-compréhension de l'homme

Pour comprendre la problématique de l'altruisme évolutionnaire, il faut connaître les principes généraux de la théorie de l'évolution. Ce chapitre est une introduction aux principes de base de cette théorie, telle qu'elle est appliquée aux niveaux biologique et culturel. Je commencerai par présenter le mécanisme de la sélection naturelle qui repose sur la valeur de la *fitness*. Je désamorcerai ensuite quelques objections courantes faites contre l'utilisation de théorie de l'évolution ; si elle est utilisée correctement, elle ne mène ni au déterminisme, ni au progressivisme, ni au réductionnisme. Il s'agira ensuite de se demander dans quelle mesure la théorie de l'évolution peut rendre compte du comportement humain ; la « thèse de la singularité humaine » sera rejetée au profit d'une compréhension de l'homme comme simple produit de l'évolution. La seconde partie de ce chapitre traitera de l'impact de l'évolution biologique sur la culture (nous verrons comment la culture a émergé et comment elle est influencée par des biais psychologiques qui résultent du processus de l'évolution) ainsi que de la meilleure manière de comprendre l'évolution culturelle elle-même.

1.1. La théorie de l'évolution biologique

1.1.1. Théorie de l'évolution et sélection naturelle

On doit la première formulation de la théorie de l'évolution des espèces biologiques à Charles DARWIN. Mais ce que l'on appelle aujourd'hui la théorie darwinienne de l'évolution incorpore un grand nombre de données que son fondateur ne pouvait ni connaître, ni même pressentir. La génétique et la biologie moléculaire, développées dès les années 1920-1930, n'étaient par exemple pas connues par DARWIN lorsqu'il publia en 1859 son fameux livre *L'origine des espèces*.

Dans l'état actuel de nos connaissances, les éléments théoriques essentiels de la

théorie de l'évolution biologique[9] sont les suivants : au cours du temps les espèces[10] changent et descendent les unes des autres par les voies ordinaires de la génération. Les espèces peuvent subsister pour une certaine période, disparaître (lorsque tous les individus qui les composent disparaissent) ou se diversifier au fil des générations (dans le cas où la composition génétique des individus d'une espèce change), donnant éventuellement naissance à de nouvelles espèces. De plus, pour toute paire d'espèces arbitrairement choisie, il a existé dans un passé plus ou moins éloigné une espèce dont elles descendent et, de proche en proche et de loin en loin, toutes les espèces peuvent être placées sur un seul grand arbre généalogique (phylogénique).

Le changement des espèces est dû à deux types de facteurs mécaniques.[11] Les premiers sont les mécanismes de diversification comme la reproduction sexuée (les divisions méiotiques et les enjambements créent de nouvelles combinaisons de gènes) ou la mutation (une erreur de transmission d'un gène due au hasard). Les seconds sont les mécanismes de tri parmi la diversité : par exemple la dérive aléatoire (à la suite de circonstances hasardeuses, tel individu survit et tel autre meurt) ou la sélection naturelle. Quoique la part de l'évolution susceptible d'être expliquée par la sélection naturelle soit l'objet de grandes controverses,[12] il s'agit du plus important mécanisme de tri parmi la diversité.

---

[9] Il existe un grand nombre d'ouvrages introductifs à la théorie de l'évolution. Je me contente de mentionner trois classiques : MAYR 1989/1982 ; J. KREBS & DAVIES 1993/1981 ; N. CAMPBELL & REECE 2004/1995 et un ouvrage récent : DAVID & SAMADI 2000.

[10] Le plus souvent, on distingue deux espèces lorsque leurs représentants ne peuvent pas se reproduire entre eux ; mais la signification exacte de ce concept reste très controversée (à ce propos, voir MAYR 1957; GAYON 2002; DAVID & SAMADI 2000, chap. XI). Une prise de position sur cette question n'étant pas primordiale pour mon propos, je ne m'y attarde pas.

[11] A ce propos, voir N. CAMPBELL & REECE 2004/1995 ; DAVID & SAMADI 2000.

[12] Contre les penseurs qui accordent une place prépondérante au mécanisme de la sélection naturelle (R. FISHER 1930 ; MAYR 1963 ; DAWKINS 1996/1976), on trouve un bon nombre d'écrits qui font la part belle à la dérive génétique (S. WRIGHT 1931 ; DOBZHANSKY 1937) ou plus généralement aux diverses forces aléatoires. Pour citer un défenseur récent de cette position : « La sélection n'est donc pas la survie systématique des plus aptes, ni même des plus féconds, comme l'écrivait Darwin, mais le tri aléatoire d'un échantillon arbitraire apte et fécond, parmi une infinité de possibles aptes et féconds. Chez l'humain comme chez le végétal ou l'animal, la génétique moléculaire des populations montre aujourd'hui que la plus grande partie de la variation observée relève effectivement de tels mécanismes aléatoires, et non d'une sélection déterministe qui ne fait qu'éliminer l'inviable et l'infécond. » (LANGANEY 2001, p. 49)

Pour qu'il y ait sélection naturelle au moins les trois conditions suivantes doivent être réunies : variation, reproduction et héritabilité (LEWONTIN 1970). Voyons ce que cela signifie dans le détail. Dans toute population, pour qu'il y ait sélection i) il faut une variation entre les traits[13] possédés par les organismes[14] composant la population considérée ; c'est-à-dire que les organismes doivent avoir des morphologies, physiologies ou comportements différents, ii) il faut que les organismes composant la population puissent survivre et se reproduire avec un taux différencié selon les traits morphologiques, physiologique ou comportementaux qu'ils possèdent, iii) il faut que les organismes parviennent à se reproduire, et par ce moyen, à transmettre leurs traits à la génération suivante.

Une autre manière de rendre compte de ce phénomène est de dire que pour qu'il y ait sélection, il faut des organismes variés du point de vue de leurs traits et de leur *fitness*[15]. Dans les écrits contemporains, on trouve différentes définitions de la *fitness*. Les acceptions les plus connues sont celles de *fitness* classique et de *fitness* inclusive.[16] Considérons pour le moment uniquement la *fitness* classique. Il s'agit d'une mesure représentative de deux facteurs : la *viabilité*, qui est la capacité d'un organisme à atteindre l'âge de reproduction et plus généralement à survivre, et la *fécondité*, qui est la capacité d'un organisme à générer une descendance.[17] On dit d'organismes dont la viabilité et/ou la fécondité sont différentes que leurs *fitness* sont différentes.[18] Ainsi, de deux individus, si au terme de leur vie, le premier est parvenu à engendrer dix petits et le second cinq, on dira du premier qu'il possédait une meilleure *fitness* que le second.

---

[13] Par « trait », j'entends une caractéristique observable d'un organisme.

[14] En réalité, le mécanisme de la sélection naturelle peut s'appliquer à autre chose qu'à un monde d'organismes vivants. Je reviendrai plus loin sur cette question (sections 1.2.2 et 1.2.3).

[15] On trouve des traductions très variées du terme *fitness* : « adaptabilité », « valeur sélective », « valeur adaptive » ou « valeur de survie et de reproduction ». La formulation anglaise sera retenue dans cet ouvrage.

[16] Il en existe toutefois d'autres. Pour un exposé plus complet, voir DAWKINS 1999/1982, chap. 10.

[17] Il est clair que du point de vue de la sélection naturelle, la viabilité n'est intéressante que dans la mesure où elle permet d'augmenter la fécondité. C'est la raison pour laquelle, en définissant le terme de *fitness*, beaucoup d'auteurs se contentent de parler du nombre de descendants sans mentionner explicitement la viabilité.

[18] « Classical *fitness* is a property of an individual organism, often expressed as the product of survival and fecundity. It is a measure of the individual's reproductive success, or its success in passing its genes on to future generations. » (DAWKINS 1999/1982, p. 182)

La *fitness* est donc une mesure relative aux individus pris en considération et ne peut être calculée qu'au terme d'un cycle complet (en l'occurrence la vie des individus dont on calcule la *fitness*).[19] Ainsi, même si l'on parle souvent de « la *fitness* d'un individu », il faut garder à l'esprit que cette propriété n'est qu'une valeur comparative ; la *fitness* n'est pas une propriété *intrinsèque* d'un organisme lui donnant, dans un environnement donné, une certaine chance de survie et de procréation. Cette dernière est sans doute mieux exprimée par la notion d'*adaptation* ; un individu adapté à son environnement a plus de chances de révéler une meilleure fitness qu'un individu moins adapté à cet environnement.

Voici une première approximation (qui sera affinée et rectifiée dans les deux sections suivantes) de la manière dont fonctionne le mécanisme de la sélection naturelle : il s'agit du processus durant lequel les organismes qui se trouvent avoir une meilleure *fitness* transmettent leurs traits, de génération en génération, dans la population globale, au détriment des organismes défavorisés du point de vue de la *fitness*. Sur le long terme, ce qui est sélectionné, ce sont des *traits.* La fréquence de certains traits au sein de la population augmentera d'autant plus que la *fitness* des porteurs de ces traits est supérieure à celle de la moyenne de la population. Voici un exemple : Si les zèbres rapides (ceux qui possèdent le trait de la rapidité) échappent plus souvent aux lions que les zèbres lents, cela implique que les zèbres rapides possèdent une *fitness* supérieure à celles des zèbres lents ; en moyenne, ces zèbres atteindront plus souvent l'âge adulte et auront une plus grande progéniture. Au fil des générations, la rapidité deviendra un trait plus fréquent dans la population des zèbres. C'est en ce sens que le processus de sélection naturelle favorise les zèbres rapides ; et en définitive, ce qui est sélectionné, c'est le trait de la rapidité. Lorsqu'un trait est sélectionné en raison

[19] Cette dernière précision permet de répondre à une série d'objections portées par DAWKINS contre l'usage de la notion de *fitness*. DAWKINS pense que la *fitness* est une notion peu utile car elle est difficile à mesurer : si elle est mesurée en termes de descendants enfants, elle ne tient pas compte du taux de mortalité infantile ; si elle est mesurée en terme de descendants aptes à avoir eux-mêmes des descendants, elle ne tient pas compte de la capacité de la deuxième génération, d'avoir des enfants aptes à se reproduire. Et si l'on prend en compte les descendants de plusieurs générations, le calcul de la *fitness* ne donnera plus que deux valeurs : zéro ou la totalité des individus de la dernière génération (DAWKINS 1999, p. 184). En réalité, on évite toutes ces difficultés si l'on s'en tient au calcul du nombre de descendants au terme de la vie des individus dont on compare la *fitness* (pour être précis, il faudrait encore, avant de les compter, attendre que les petits aient atteint l'âge adulte).

de l'avantage qu'il procure à l'organisme porteur, on peut le considérer comme une « adaptation ».

Deux remarques importantes s'imposent ici. Premièrement, la sélection naturelle ne modifie pas les traits des individus eux-mêmes ; ce qu'elle modifie au fil des générations, ce sont les proportions dans lesquelles ces traits sont présents dans les individus de la population en général. Deuxièmement, la sélection naturelle opère toujours dans le cadre d'un *environnement* donné ; si l'environnement change, des traits peuvent perdre leur valeur adaptative et d'autres traits seront sélectionnés.

1.1.2. La perspective du gène

Depuis le début du vingtième siècle, grâce à la redécouverte (DE VRIES 1900) des travaux de MENDEL (1911/1865), on admet que les traits héréditaires ne se transmettent pas eux-mêmes, mais par l'intermédiaire de gènes. Prenant conscience de cette base génétique des traits et comportements observables, les théoriciens de l'évolution ont pris l'habitude de penser en termes de gènes et de leurs phénotypes.[20] Un « phénotype » est l'effet perceptible d'un ou plusieurs gènes ; il s'agit, soit d'un trait physique, soit d'une tendance à agir d'une certaine manière dans certaines circonstances.[21]

Cette nouvelle manière de concevoir la transmission de caractères héréditaires en termes de relation gène-phénotype a également eu pour conséquence de minimiser le rôle de l'individu dans le processus de sélection naturelle en le reléguant à la simple fonction de « véhicule » ou « machine à transporter les gènes ».[22] C'est depuis lors que

[20] La notion de « phénotype » tout comme celle de « gène » et de « génotype » sont dues à Wilhelm JOHANNSEN (1909 ; 1911). Pour un exposé exemplaire des débuts de la génétique et de ses implications, voir GAYON 2000.

[21] Les gènes n'ont pas uniquement des effets phénotypiques sur les corps et les comportements des individus dans lesquels ils se trouvent ; ils peuvent aussi avoir des effets phénotypiques *étendus*, c'est-à-dire qui touchent des objets ou des individus extérieurs aux individus porteurs des gènes en question. Par exemple, certains gènes des castors ont pour effet phénotypique étendu la création de petits lacs en amont des barrages que les castors construisent dans les rivières. Ou alors, les oisillons qui piaillent très fort ont pour effet phénotypique d'inciter leurs parents à les gaver avant leurs frères moins bruyants (car le piaillement implique le risque que le nid soit détecté par un prédateur). A ce propos, voir DAWKINS 1996/1976, pp. 318-339 et DAWKINS 1999/1982.

[22] La formule a été rendue célèbre par Richard DAWKINS (1996/1976) mais l'instigateur de cette approche « génique » est Ronald FISHER (1930).

l'on a commencé à adopter la perspective du gène, c'est-à-dire à penser que la sélection naturelle opère au niveau des gènes plutôt qu'au niveau des individus et de leurs phénotypes (R. FISHER 1930 ; HAMILTON 1964 ; G. WILLIAMS 1966 ; MAYNARD SMITH & G. PRICE 1973 ; DAWKINS 1996/1976). Ce qui importe désormais, c'est la manière dont un gène responsable d'un phénotype peut, par le biais de la réplication (création de copies exactes de lui-même), se répandre dans l'ensemble du pool génétique d'une population. Selon cette perspective, seuls les gènes capables de se répliquer plus que les autres sont favorisés par la sélection et la *fitness* d'un individu ne compte que dans la mesure où elle sert « l'intérêt »[23] des gènes véhiculés par cet individu.[24]

Ainsi, pour revenir à l'exemple du zèbre, le trait de la rapidité est un phénotype, c'est-à-dire l'expression d'un gène. Sachant cela, en un sens on pourrait penser que c'est le gène (en tant que type) plus que le trait de la rapidité, qui a été sélectionné, la rapidité n'étant que l'expression de ce gène. Il y a toutefois deux raisons de se méfier d'une conception trop matérialiste du processus de sélection naturelle.

Premièrement, la plupart des gènes (il y a des exceptions) ne peuvent être sélectionnés que si leur expression est favorable aux individus qui les portent. En ce sens, on peut dire que la sélection biologique opère plutôt sur les fonctions[25] que sur les structures. Ce n'est qu'indirectement, par le biais des fonctions, que les structures sous-jacentes sont sélectionnées. Je m'explique. Les objets que considèrent les théoriciens de l'évolution (les traits physiques et comportementaux) ont à la fois une structure et une

[23] La notion d'« intérêt » des gènes doit être comprise dans un sens particulier (au même titre que la notion de « gène égoïste » qui est utilisée par certains auteurs de manière purement métaphorique). D'une part les gènes n'ont pas d'intentions de sorte qu'ils ne peuvent pas chercher à maximiser leurs intérêts ; il n'y a donc pas lieu de parler d'*intérêt subjectif* pour les gènes. D'autre part, du point de vue évolutionnaire, les gènes ne sont intéressants qu'en tant que types si bien que l'on ne peut même pas parler d'*intérêt objectif* pour *une* occurrence matérielle d'un gène (un brin d'ADN). A ce propos, voir MACKIE 1989.

[24] Dès lors, la *fitness* classique qui est une manière de calculer les avantages sélectifs en tenant uniquement compte du succès reproductif des individus (c'est-à-dire leur capacité de transmettre leurs propres traits à leurs enfants, petits enfants, etc.) devient une mesure moins intéressante. Elle sera remplacée par le fameux calcul de la *fitness* inclusive développé par William HAMILTON (1964). Nous aurons l'occasion de revenir sur ce point.

[25] Dans ce contexte, je m'inspire de l'approche fonctionnaliste étiologique de fonction (L. WRIGHT 1973). Pour plus de détails sur les différentes manières de concevoir la notion de fonction, voir L. WRIGHT 1973 ; CUMMINS 1975 ; PROUST 1995.

fonction biologique (ou plusieurs). La structure, c'est ce dont est composé l'objet ; ses caractéristiques physiques particulières, ses bases génétiques, etc. La fonction, de manière très schématique, c'est l'effet de la structure sur le monde. Plus précisément, *X* est la fonction biologique de *y* si le fait que *y* a pour effet *X* est le résultat d'une sélection au fil de l'évolution. Illustrons cela par deux exemples. Prenons un nez. Une de ses fonctions biologiques est le fait d'inspirer de l'oxygène. En revanche, le fait d'être un support de lunettes n'est pas la fonction biologique du nez ; le nez n'a pas été sélectionné parce qu'il peut servir de support de lunettes. Voici un second exemple. Considérons un comportement de fuite face à l'arrivée d'un prédateur : chez les lapins une fonction biologique évidente de ce type de comportement est l'augmentation des chances d'échapper aux griffes des prédateurs. Ce type de comportement a été sélectionné parce que sa fonction s'est avérée bénéfique à la survie des lapins.

Deuxièmement, Elliott SOBER (1984a ; 1984b) a introduit une distinction importante qui incite à ne pas focaliser uniquement l'attention sur la transmission génétique lorsque l'on analyse les processus de l'évolution. Il distingue entre « sélection de », qui se réfère uniquement à ce qui est effectivement sélectionné (en l'occurrence le gène responsable du trait de la rapidité) et « sélection pour », qui se réfère au fait qu'un gène a été sélectionné en raison de l'avantage produit au niveau de l'organisme. Le terme de « sélection pour » souligne bien le fait que le mécanisme de la sélection naturelle ne porte pas uniquement sur des réplicateurs mais également sur les porteurs de ces réplicateurs, ceux qui interagissent réellement dans un environnement (à ce propos, voir HULL 1980 ; GAYON 1999).

1.1.3. Trois faux problèmes : progressivisme, réductionnisme, déterminisme

La théorie de l'évolution se caractérise à la fois par le hasard et par la nécessité : le hasard intervient par le biais des mutations et de la dérive aléatoire ; la nécessité intervient dans le processus de la sélection naturelle (ce sont généralement les individus les mieux adaptés qui survivent). Ainsi, d'un côté l'évolution se caractérise par un facteur non déterministe (le hasard), de l'autre elle est déterminante dans le sens où chaque pas évolutif est soumis aux contraintes de la sélection naturelle qui en

conditionne le développement.[26]

Puisqu'il y a nécessité, on pourrait se demander si l'évolution se dirige vers le plus parfait, vers un monde *meilleur* ; finalement, ce sont toujours les traits les plus adaptés qui sont sélectionnés... A cette question, il faut répondre par la négative car toute adaptation par sélection naturelle est relative, non absolue. Un trait phénotypique (par exemple un pelage épais) peut être extrêmement bien adapté à un type d'environnement, mais pour peu que cet environnement se transforme (par exemple un brusque réchauffement climatique), il ne sera plus adapté.[27] Ou alors ce qui est optimal selon une perspective ne l'est pas forcément selon une autre. Par exemple, un trait avantageux pour la reproduction ne l'est pas forcément pour la survie ;[28] il se peut qu'un phénotype donné accroisse la fécondité d'un organisme tout en réduisant sa viabilité. L'exemple le plus connu est celui des longues plumes colorées du paon : l'animal en tire un avantage du point de vue de la fécondité (ses plumes plaisent aux femelles), par contre il est désavantagé du point de vue de la viabilité (les couleurs vives attirent l'œil du prédateur et la longueur des plumes entrave la fuite).[29] Ainsi, la sélection naturelle sélectionne moins les plus aptes à survivre dans un environnement donné (en l'occurrence les plus forts) que les plus aptes à s'y reproduire. Enfin, comme nous le verrons plus loin dans les discussions sur l'altruisme, un trait peut s'avérer adaptatif à un niveau de sélection mais non à un autre.

---

[26] Notons en passant que ce facteur de hasard empêche toute interprétation finaliste de l'évolution ; l'évolution ne poursuit aucun dessein (à ce propos, voir DENNETT 2000/1995, première partie).

[27] Ce phénomène est notamment pris en considération par le généticien des populations Sewall WRIGHT, lorsqu'il présente le conflit permanent entre les optima locaux (du point de vue de la *fitness*) et les optima globaux d'une lignée. Dans ce contexte, S. WRIGHT montre qu'à partir du moment où un optimum local est atteint, il demeure très fragile et sensible au moindre changement de l'environnement (S. WRIGHT 1932).

[28] Citons Ernt MAYR à ce propos : « La sélection ne peut produire la perfection, car dans la compétition pour le succès reproductif entre les membres d'une population, il suffit d'être supérieur, il n'est pas nécessaire d'être parfait. » (MAYR 1989/1982, p. 545).

[29] Selon la fameuse théorie du handicap développée par Amotz ZAHAVI (1975), le coût évolutionnaire engendré par certains handicaps (tels que les longues plumes de certains oiseaux) serait un signal fiable pour les femelles de la bonne santé des mâles (puisqu'ils parviennent à survivre malgré ce handicap), si bien que les femelles préfèreraient les mâles présentant ces handicaps (à ce propos, voir section 2.2.3).

Une autre question surgit en rapport avec le caractère nécessaire de l'évolution. Il s'agit de savoir si les nouveaux et futurs acquis en matière de génétique auront pour conséquence de permettre à la théorie de l'évolution de donner une explication *réductionniste* (en termes de code génétique) et *déterministe* de tous les phénomènes vivants observables. Le physicien Erwin SCHRÖDINGER (1993/1994) par exemple, a écrit un petit ouvrage très lu dans lequel il défendait l'idée que chaque cellule d'un être vivant contient l'information génétique qui programme son existence.

On sait aujourd'hui que, pour présenter les choses de manière très grossière, un morceau d'ADN (molécule formée de nucléotides) est transcrit en un morceau d'ARN (molécule également formée de nucléotides), lequel est traduit sous forme de protéines (formées d'acides aminés), lesquelles composent les corps. Peut-on en déduire une relation biunivoque entre un gène (une séquence d'ADN codant) et une séquence de la protéine codée par ce gène ? La réponse est négative. Ces dernières années, l'avancée des recherches montre à quel point une telle approche est simpliste. Suite à de nombreuses recherches en biologie moléculaire, force est de constater que la correspondance biunivoque entre gène et protéine n'est pas réalisée. Voici quelques raisons de rejeter cette relation biunivoque. On a constaté qu'un seul gène peut produire plusieurs protéines différentes (RIDLEY 2004/2003, pp. 137-144) et inversement, une même protéine peut être le résultat de l'interaction de plusieurs gènes. Il semblerait également que 80 à 90% de l'ADN soit inutile (on n'en a du moins pas encore décelé l'utilité) ; seul 10 à 20% de l'ADN génomique global peut être considéré comme matériel génétique (DAVID & SAMADI 2000, pp. 49-52). D'autre part, les généticiens ont pu constater que la transcription de l'information génétique est bien plus complexe que ce qu'on avait d'abord imaginé. Aujourd'hui, on sait par exemple que les ARN ne se limitent pas à assurer le transit de l'information des gènes aux protéines ; certains types d'ARN semblent avoir de multiples fonctions en agissant par exemple comme répresseurs génétiques ou inhibiteurs de traduction (MATZKE & KOOTER 2001). On sait également que l'environnement cellulaire (une cellule n'est pas uniquement faite d'ADN puisqu'elle contient un cytoplasme composé de protéines, elles-mêmes codées par d'autres gènes) joue un rôle dans le processus de transcription des gènes pour la formation de protéines ; il peut par exemple empêcher la traduction puis la transcription d'un gène (MAYNARD SMITH 2001/1998 ; KUPIEC & SONIGO 2000 ; PENNISIS 2001).

Toutes ces nouvelles découvertes nous poussent à réviser notre conception du couple gène/phénotype. On doit admettre qu'un gène peut être responsable de plusieurs

effets phénotypiques (par exemple, un même gène peut à la fois être responsable des yeux bleus et des cheveux frisés) et, inversement, qu'un effet phénotypique peut être causé par différents gènes (par exemple, il est possible qu'il existe plusieurs centaines de gènes capables d'induire la création d'une épaisse fourrure ; voir RIDLEY 2004/2003, pp. 129-130). On doit également considérer que l'environnement dans lequel l'individu évolue exerce une très grande influence sur l'expression (ou réalisation) des effets phénotypiques. Ainsi, à partir d'un même génotype, plusieurs phénotypes sont possibles, et c'est l'environnement qui oriente l'organisme vers un phénotype particulier plutôt que vers un autre (dans ce contexte, on parle d'influences « épigénétiques ») ; par exemple, le même papillon, selon la flore dans laquelle il grandit, peut développer des ailes de couleurs différentes. C'est ce qu'on appelle le phénomène de la *plasticité phénotypique* (DAVID & SAMADI 2000, chap. 9).[30]

En définitive, l'ADN n'est pas porteur d'un programme fait d'instructions rigides, dans lequel l'organisme adulte serait écrit à l'avance, si bien que les biologistes et généticiens sont réduits à s'exprimer en termes de probabilités qu'un génotype s'exprime d'une certaine manière phénotypique ; ce qu'ils peuvent affirmer, c'est qu'il existe une corrélation (plus précisément une causalité statistique) entre un gène et l'occurrence d'un effet phénotypique.

Faut-il en conclure que cette difficulté à comprendre la relation entre les gènes et leurs phénotypes remet en cause la théorie de la sélection naturelle elle-même et plus généralement la théorie de l'évolution ? Je ne le pense pas.

D'une part, la force de la théorie de l'évolution se mesure à l'étendue des données qu'elle se montre capable d'intégrer et à sa capacité d'expliquer les phénomènes. Or, il se trouve qu'elle est aujourd'hui la seule théorie scientifique de l'évolution des vivants (cela n'empêche pas qu'il reste beaucoup d'énigmes à résoudre). Pour se permettre de rejeter cette théorie il faudrait en proposer une autre dont la force explicative est au moins aussi grande que celle de la théorie de l'évolution.[31]

[30] En réalité cette plasticité phénotypique est très avantageuse pour les gènes puisqu'elle permet à l'individu porteur de s'adapter à diverses conditions environnementales. De manière un peu sommaire, on peut même ajouter que le déterminisme n'est pas dans l'intérêt des gènes.

[31] Cette question est étroitement liée aux débats qui font actuellement rage entre les théoriciens de l'évolution et les défenseurs du « dessein intelligent » (*intelligent design*). Pour une critique circonstanciée de la position de ces derniers, voir SOBER 2007.

D'autre part, je ne pense pas qu'une prise de position sur ces questions soit indispensable à la théorie de la sélection naturelle. Rappelons-nous (p. 25) que si l'on échappe à la perspective matérialiste du gène, on comprend que le plus souvent, la sélection biologique opère directement sur la fonction et indirectement (par le biais de la fonction) sur la structure. Ainsi la connaissance de cette dernière n'est pas absolument nécessaire à l'application de la théorie de l'évolution.

1.1.4. La théorie de l'évolution appliquée à l'homme : rejet de la thèse de la singularité humaine

La théorie de l'évolution ne prétend pas uniquement expliquer les traits observables ; elle cherche également à rendre compte des comportements des individus. Lorsqu'on cherche à expliquer des comportements génétiquement déterminés (chez les insectes par exemple), les choses sont encore relativement simples ; on peut assez bien observer leur évolution puisqu'ils sont rigides ; ce sont des objets de sélection qui sont transmis de manière stable au fil d'un grand nombre de générations. Tout se complique lorsqu'on considère les comportements sociaux des mammifères (singes, lions, dauphins, etc.). En effet, ceux-ci témoignent d'une certaine finesse cognitive, voire d'un apprentissage culturel qui échappe partiellement aux déterminations génétiques (DE WAAL 1997/1996).[32]

Les difficultés décuplent lorsqu'il s'agit d'expliquer les comportements propres aux êtres humains. On sait combien l'espèce humaine se différencie des autres espèces biologiques connues, par la complexité de son monde culturel, sa grande capacité d'apprentissage, sa maîtrise du langage, sa capacité d'effectuer des choix, son incroyable faculté de réflexion, ses qualités en matière de coopération et de répartition du travail et surtout sa manie de la moralisation.

Dès lors se pose la question de savoir s'il est encore pertinent d'utiliser les méthodes de la théorie de l'évolution pour expliquer le comportement humain. On pourrait répondre par la négative en affirmant que grâce à ses facultés inédites, l'homme

[32] Si l'on adopte une définition minimale de l'apprentissage culturel au sens où il y a information transmise par le biais d'interactions sociales, alors on peut parler de culture chez les animaux. Nous verrons plus loin que la culture animale est généralement extrêmement rudimentaire par rapport à la culture humaine. Elle se résume plus ou moins à la capacité d'imiter le chant de son voisin ou au phénomène de renforcement local (pour ce dernier, voir section 1.2.1, p. 34).

peut disposer librement de lui-même et échapper dans une large mesure aux contraintes du monde biologique. Ce serait défendre « la thèse de la singularité humaine ».[33] Selon cette thèse, les actions humaines relèveraient du domaine culturel, lequel est régi par une causalité particulière et indépendante de l'influence naturelle (TORT 2002).

Pour commencer, il convient de remarquer que même à supposer que cette thèse soit juste, il est envisageable que la causalité propre au domaine culturel soit elle-même régie par un mécanisme de sélection naturelle ; en effet, nous verrons plus loin (sections 1.2.2 et 1.2.3) que l'on peut concevoir un phénomène d'évolution culturelle qui fonctionne plus ou moins sur le modèle de la sélection naturelle. On ne se débarrasse donc pas si aisément de l'approche évolutionnaire.

Cela dit, je pense que la thèse de la singularité humaine doit être rejetée d'une part parce qu'elle incite à adopter des thèses spécieuses, d'autre part parce qu'elle se prive d'une importante dimension explicative du comportement et de la pensée humaine.

Pour ce qui est du premier point, il me semble que la thèse de la singularité humaine est liée (même s'il n'y a pas à proprement parler d'implication logique) à la tentation d'attribuer un statut d'exception à l'être humain. Or il vaudrait mieux être prudent sur ce chapitre. Dire de l'homme qu'il est particulier parce qu'il possède des capacités que l'on ne peut pas trouver chez les animaux (l'intentionalité de deuxième ordre ou le langage par exemple) n'est pas problématique à condition que l'on ne voie aucune différence essentielle, au sens où l'homme serait supérieur aux autres espèces. Au fond on trouve aussi chez les autres espèces des facultés dont nous ne pouvons que rêver (voir dans le noir, voler, etc.). Ainsi Joelle PROUST écrit très justement :

> « Aucun animal non humain ne brille dans ces différentes formes de raisonnement – si précieuses à nos yeux précisément parce qu'elles forment les conditions de l'apprentissage dont dépend la transmission de la culture humaine. Inutile de tirer gloire de cette culture humaine ou d'en vanter la supériorité. On ne pourrait dire que l'homme est supérieur à l'animal que si chaque espèce avait eu le même environnement et les mêmes problèmes à résoudre. S'il est une leçon que l'on peut tirer de la biologie, c'est que le cerveau et les capacités représentationnelles d'une espèce donnée lui ont permis de résoudre les problèmes particuliers qui se sont présentés dans son passé. Il n'est donc pas possible de tirer des conclusions normatives de la comparaison entre les manières de représenter le monde. » (PROUST 2003, pp. 158-159)

[33] Pour une critique de cette thèse, voir MACHERY 2003.

Il est vrai que du point de vue de l'évolution, très peu de temps s'est écoulé entre le passage à l'agriculture et nos jours. Cela laisse présumer qu'aucune évolution génétique majeure n'a pu affecter notre patrimoine génétique de manière significative dans ce laps de temps.[34] Il s'ensuit que l'on ne dispose pas de capacités génétiquement adaptées à bon nombre de caractéristiques environnementales propres au monde contemporain (l'usage de l'écriture, les différentes communautés langagières, l'avancée de la technologie, etc.). De plus, à n'en pas douter, nous avons passablement changé dans nos habitudes de vie et acquis une gamme de nouvelles connaissances. Mais ces réalités ne peuvent pas être utilisées par les défenseurs de la thèse de la singularité humaine comme preuve de l'indépendance humaine par rapport à sa nature biologique. Il faut distinguer entre le contenu ou l'étendue des connaissances (coutumes, techniques, etc.) et les capacités qui nous permettent d'acquérir ces connaissances. Or, comme nous le verrons dans les sections suivantes, y a de bonnes raisons de penser que notre cerveau est composé de nombreux sous-mécanismes qui sont apparus comme des adaptations aux contraintes environnementales auxquelles étaient confrontés nos ancêtres (voir section 1.2.3). Ces sous-mécanismes (dispositions au comportement social, mécanismes de traitement des données reçues, etc.) nous permettent d'acquérir des représentations mentales et en influencent encore aujourd'hui le contenu. Adopter la thèse de la singularité humaine nous prive de ce type d'explications très éclairantes sur le comportement humain. Ainsi, je suis d'avis que si l'on veut comprendre le comportement et la pensée humaine, au lieu de faire de l'homme un objet singulier, il est plus utile de le placer dans le contexte de l'évolution et de tâcher de déterminer les capacités qu'il a acquises en réponse aux milieux dans lesquels il a évolué. Comme le remarque Dominique LESTEL, il faut comprendre que « l'homme n'est pas sorti de l'état de nature, mais il en a exploré avec succès une niche extrême » (LESTEL 2003/2001, p. 162). En résumé, beaucoup de particularités propres à l'homme sont dues au formidable développement de ses capacités cognitives ; ces dernières peuvent être comprises comme des réponses adaptatives aux nécessités de la survie de l'espèce humaine.

[34] « We can be fairly sure that the nature (i.e. the genetically determined capacities) of human beings has not greatly changed since the Neolithic revolution, since 7,000 years is too short a period for major evolutionary changes. » (MAYNARD SMITH 1993, p. 328)

1.2. Evolution et culture humaine

Le rapport entre l'évolution biologique et la culture est extrêmement complexe. Je vais tâcher de l'éclairer quelque peu dans les prochaines sections en défendant l'idée que l'évolution biologique exerce un impact non négligeable sur la culture : d'une part, elle rend la culture possible, d'autre part elle en influence le contenu. Les explications qui seront proposées ne permettront certainement pas de décider la part de l'influence de nos gènes et celle de la culture pour un comportement particulier. Elles mettront par contre en évidence les capacités sous-jacentes à la culture et ouvriront la possibilité de découvrir des biais psychologiques[35] auxquels il nous est difficile de résister et qui structurent notre pensée et nos comportements.

Cela dit, il est clair que la détermination génétique ne permet pas d'expliquer le phénomène culturel dans son entier. Il existe une telle variation culturelle et les changements s'opèrent à un rythme si effréné à l'échelle historique que le « recours au gène » n'est d'aucune utilité pour comprendre le processus de création des différents langages, habitudes de comportement, codes sociaux, etc. Pour comprendre les modifications culturelles rapides, nous verrons qu'il faudra recourir à des schèmes explicatifs propres à l'évolution culturelle ; je présenterai deux manières de concevoir l'évolution culturelle, dont seule la seconde est viable.

1.2.1. L'émergence de la culture

Si l'on comprend le mot « culture » au sens où il y a transmission, acquisition et accumulation d'information[36] entre individus (HENRICH & MCELREATH 2003, p. 124),[37]

[35] Précisons que la notion de biais psychologique est utilisée ici au sens de tendance psychologique sans qu'aucune connotation négative n'y soit associée ; plus de détails sur cette notion seront donnés à la section 1.2.3.

[36] La notion d'information doit être comprise dans un sens large ; elle se rapporte en principe à tout état mental, conscient ou non, qui est acquis ou modifié par le biais de l'apprentissage social (RICHERSON & BOYD 2005, p. 5). Ces états mentaux transmis d'un esprit à l'autre ont évidemment des conséquences au niveau du comportement des individus.

[37] Notons que cette conception de la culture est propre aux approches évolutionnaires. Elle saisit le phénomène de la culture en termes quantitatifs, en termes de transmission d'entités culturelles. Comme le fait bien remarquer Elliott SOBER (1994b/1993, pp. 488-489), en sciences sociales par exemple, on s'intéresse plutôt à l'aspect qualitatif des phénomènes. Dan SPERBER et Nicolas CLAIDIÈRE (2008) mettent

alors il semblerait que le mécanisme qui a rendu possible la culture (c'est-à-dire qui a permis aux être humains de recevoir et transmettre de l'information) soit l'*imitation*, qui est une forme d'apprentissage par observation (TOMASELLO *et al.* 1993 ; TOMASELLO 2004/1999). Evidemment, d'autres formes de transmission sociale, plus complexes, comme l'enseignement, le langage ou l'écriture favorisent grandement la transmission culturelle et surtout l'accumulation d'information ; mais l'imitation semble jouer le rôle clé de *condition nécessaire* de la culture telle qu'elle est définie ci-dessus (d'autant plus que les formes de transmission plus élaborées dépendent de la capacité d'imiter).

Les recherches en éthologie et plus particulièrement en primatologie semblent montrer que l'imitation est très rare dans le monde animal à l'exception peut-être de certaines espèces d'oiseaux ou de grands singes (ZENTALL 2006). Il existe en revanche toute panoplie d'autres formes moins efficaces d'apprentissage social, que l'on trouve dans le monde animal et qui permettent également de parler d'hérédité culturelle ou de culture *rudimentaire*. Par exemple, ce que l'on appelle le « renforcement local » est un phénomène social qui consiste en une haute probabilité que les individus apprennent une technique par eux-mêmes parce qu'ils sont exposés à toutes les conditions qui favorisent l'acquisition de cette technique. En voici un exemple. Supposons qu'un animal découvre par hasard une technique qui lui permet d'obtenir une nouvelle ressource et l'utilise ensuite régulièrement. Dans certaines régions d'Angleterre, par exemple, des mésanges ont découvert la technique du décapsulage des bouteilles de lait déposées devant les maisons afin d'en boire la crème (J. FISHER & HINDE 1949). S'il s'agit d'une espèce un tant soit peu grégaire, d'autres individus seront présents lorsque l'inventeur utilise sa nouvelle technique ; de ce fait, ils seront donc globalement soumis aux mêmes stimuli et finiront par découvrir par eux-mêmes comment obtenir cette ressource. Les exemples de renforcement local les plus connus peuvent être trouvés chez les singes : dans certaines régions, les macaques apprennent à laver certains de leurs aliments (les patates douces) avant de les manger et les chimpanzés apprennent à attraper les termites dont ils raffolent en plantant une baguette dans la termitière et en attendant patiemment qu'elles s'y agrippent (BOESCH & TOMASELLO 1998). Il est

également en garde contre les simplifications excessives du phénomène de la culture. Je pense que ces auteurs ont raison de souligner le fait que les théories évolutionniste ne peuvent saisir qu'un aspect (aussi fondamental soit-il) de la culture ; cela n'invalide toutefois en rien l'intérêt de l'approche évolutionnaire.

important de noter ici que dans le cas du renforcement local, les nouvelles techniques ne sont pas transmises directement par observation et imitation des agissements d'autres individus. En revanche, il y a une composante sociale dans cet apprentissage, puisqu'il implique un attroupement d'individus sur le même site. Le renforcement local est une méthode d'apprentissage social moins efficace que l'imitation ; puisque chaque individu par lui-même doit réinventer les détails du comportement, ce dernier ne peut pas devenir plus complexe au fil des générations. Le renforcement local permet donc de maintenir des traditions mais non d'accumuler des connaissances ou des innovations.

En plus du renforcement local et de l'imitation, il existe tout une série d'autres mécanismes comme l'émulation, l'amélioration du stimulus (stimulus enhancement), etc. (voir BOESCH 1996 ; ZENTALL 2006). Mais pour notre propos, l'imitation est plus intéressante que les autres formes d'apprentissage social dans la mesure où elle permet d'acquérir de nouveaux comportements directement par le moyen de l'observation et de la reproduction détaillée des agissements d'autres individus. L'imitation, cette faculté de copier les acquis du travail d'apprentissage effectué par d'autres individus, permet d'intégrer les innovations précédentes ; elle est à la source de la culture humaine dans toute sa complexité (R. BOYD & RICHERSON 1985 ; TOMASELLO 2004/1999).

Les découvertes archéologiques (taille du cerveau humain corrélé à la production d'outils ou nouvelles techniques et aux pratiques de la parure et de l'enterrement des morts) portent à croire que l'acquisition de la capacité d'imitation remonte à 40'000 ou 100'000 ans. Il semblerait que durant la première période de l'évolution des hommes, l'apprentissage par observation n'était pas très développé et devait plutôt ressembler au renforcement local. « L'outil le plus perfectionné utilisé par H. erectus était un genre de hachoir, formé d'un seul bloc de pierre travaillé sur deux surfaces et de forme symétrique. Les premiers hachoirs sont apparus il y a 1,4 millions d'années et sont restés presque inchangés pendant un million d'années : ce n'est pas vraiment un exemple d'accumulation de changement culturel ! » (MAYNARD SMITH & SZATHMARY 2000/1999, p.162). Ce n'est probablement pas avant 100'000 ou 40'000 ans av. J.-C qu'est apparu l'homme moderne capable d'acquérir et de transmettre des informations par le biais de l'observation et de l'imitation.[38] L'accumulation des découvertes

[38] Ces dates sont cependant sujettes à controverse. Comme le font remarquer les anthropologues Sally MCBREATY et Alison BROOK (2000), il n'y a probablement pas eu de « révolution culturelle » et l'*homo*

archéologiques d'objets divers datant du paléolithique supérieur parle en ce sens.[39]

Si la culture est apparue, il est hautement probable qu'elle s'avère avantageuse du point de vue évolutionnaire. C'est du moins ce que pensent un bon nombre d'auteurs (BOYD & RICHERSON 1985 ; LUMSDEN & E. WILSON 1981). Selon eux, l'apprentissage social s'avère moins coûteux que l'apprentissage individuel et donne accès à un corpus d'informations (techniques, pratiques, connaissances) qu'il ne serait pas possible d'assimiler par soi-même au cours d'une seule vie. Ainsi, la possibilité d'accumuler de l'information par le biais de l'imitation sans passer par la phase d'apprentissage individuel comporte un avantage sélectif indéniable ; dans bien des contextes, il s'avère plus avantageux de copier les idées et comportements avantageux plutôt que de perdre temps et énergie à apprendre par soi-même au moyen de la pratique de l'essai et de l'erreur. L'avantage direct procuré par la pratique de l'imitation a permis la sélection de cette capacité.

Mais l'imitation a également un coût non négligeable : produire des capacités cognitives nécessaires à l'imitation.[40] Ainsi, pour qu'un tel investissement soit rentable du point de vue évolutionnaire, il faut que l'apprentissage de nouveaux comportements et techniques soit avantageux par rapport au fait de perpétuer les habitudes ancestrales. Cela ne peut être le cas que dans des environnements changeants auxquels les individus doivent constamment se réadapter. Or il semblerait que ce soit précisément le cas : depuis le pléistocène, notre terre a subi de grandes variations climatiques. Ainsi tout

---

sapiens a acquis peu à peu depuis plus de 200'000 ans, les capacités que nous lui connaissons aujourd'hui. « The complex content of human cultures has been built incrementally, with cognitive equipment present since at least 250 ka, in a process that continues today » (2000, pp. 531-532)

[39] « L'accroissement de la taille du cerveau s'est accéléré ces derniers 300'000 ans pour culminer avec l'apparition de l'homme moderne il y a quelques 100'000 ans. Cependant, l'accélération de l'inventivité technique humaine, marquée par l'apparition d'une gamme d'outils de pierre, de bois et d'ivoire remonte à 40'000-50'000 ans. L'enterrement des morts, l'art rupestre, les instruments de musique, les parures et le commerce datent à peu près de la même époque. » (MAYNARD SMITH & SZATHMARY 2000/1999 p. 164)

[40] Certains auteurs (HENRICH & MCELREATH 2003) pensent par exemple que l'imitation est dépendante de l'évolution d'une capacité cognitive encore plus fondamentale : la « théorie de l'esprit » (*Theory of Mind* ou *ToM*). Il s'agit de la capacité de raisonner au sujet des états mentaux d'autrui (pour plus de détails, voir section 3.1, p. 137). Avoir une théorie de l'esprit permet de comprendre les intentions d'autres individus, ce qui est un facteur essentiel pour l'imitation. Selon HENRICH et MCELREATH, la culture ne serait même qu'un épiphénomène de l'évolution de la théorie de l'esprit.

porte à penser que la culture (et avec elle l'accroissement des capacités cognitives) est une excellente réponse adaptative au fait que les êtres humains ont dû survivre dans des environnements instables (SOLTIS *et al.* 1995 ; RICHERSON & BOYD 2000).

Alan ROGERS (1988) a émis quelques critiques à l'encontre de cette explication de l'évolution de la culture (en particulier la version de BOYD ET RICHERSON 1985). Selon lui, épargner aux individus les coûts de l'apprentissage individuel n'est pas suffisant pour augmenter l'adaptabilité moyenne d'une population. Sans individus capables d'apprendre par eux-mêmes, la population ne peut plus s'adapter à des changements de l'environnement et les avantages de l'apprentissage social chutent de manière dramatique. En réponse, Robert BOYD et Peter RICHERSON (SOLTIS, BOYD & RICHERSON 1995) ont proposé un modèle mixte où les êtres humains sont capables de changer de stratégie selon les situations : apprendre par eux-mêmes quand les informations sont accessibles à peu de frais et les chances de réussite assez grandes, sinon, copier soit les idées et stratégies comportementales qui semblent apporter les meilleurs résultats, ou bien les idées et stratégies utilisées par les individus dont la position sociale est la plus enviable. Combiné à l'apprentissage individuel, l'apprentissage social est avantageux car il permet d'assimiler rapidement de nouvelles connaissances et techniques adaptées ; cela a permis à nos ancêtres de gérer les importants changements environnementaux auxquels ils ont dû faire face.[41]

1.2.2. Evolution culturelle et évolution biologique: une analogie stricte

Dire que la culture est un produit de l'évolution n'impose pas d'adopter une approche qui en minimise le rôle en faisant de toute entité culturelle le résultat d'une sélection biologique. Cette position a pourtant été défendue par certains auteurs comme Mark FLINN et Richard ALEXANDER (1982 ; voir aussi IRONS 1979) ; s'ils admettent que les productions culturelles ne sont pas simplement des expressions phénotypiques de

[41] Je n'ai présenté ici qu'une partie des explications de l'évolution de la culture. La littérature sur ce sujet est si vaste que je ne peux en esquisser ici que quelques grandes lignes. Beaucoup de questions restent encore ouvertes, notamment celle de savoir comment une machinerie cognitive complexe et coûteuse comme l'imitation a pu apparaître en premier lieu et atteindre le seuil critique à partir duquel elle s'est avérée réellement efficace et évolutionnairement stable (concernant cette question, voir BOYD & RICHERSON 1996). Pour une revue de la littérature et de nouveaux développements, voir HENRICH & MCELREATH 2003 ; ALVARD 2003; STERELNY 2006.

gènes, ils soutiennent en revanche qu'elles peuvent uniquement être sélectionnées si elles favorisent la *fitness* biologique des individus producteurs de ces caractéristiques. Ainsi les productions culturelles sont sélectionnées exactement de la même manière que les phénotypes des gènes, en fonction des avantages qu'elles apportent aux individus qui les pratiquent. Ce modèle est cependant assez peu convaincant précisément parce qu'en focalisant sur les résultats en termes de survie des organismes transmetteurs d'entités culturelles il est insensible à la dynamique culturelle elle-même. Ce modèle passe à côté de la complexité des processus culturels en ignorant une réalité indéniable : la transmission des entités culturelles procède en bonne partie de manière indépendante de leurs effets sur la survie des organismes.

Contrairement à ce genre d'approches, il existe des modèles selon lesquels la sélection des entités culturelles ne relève pas d'un processus biologique mais se fait plutôt sur le mode de la propagation d'un virus dans une population. Il s'agit d'un système autonome qui fonctionne de la même manière que la sélection naturelle biologique ; mais alors que là, ce sont les gènes et leurs phénotypes qui sont transmis et sélectionnés, ici ce sont des entités culturelles. Ainsi on constate que la théorie de l'évolution ne se cantonne pas au monde strictement biologique ; elle peut être appliquée dans différents contextes.

La théorie de l'évolution culturelle classique a été proposée pour la première fois par Richard DAWKINS (1996/1976). Elle a séduit un certain nombre d'auteurs (DENNETT 2000/1995) et fait l'objet de nouveaux développements (BLACKMORE 1999). Elle repose sur une analogie stricte avec l'évolution biologique et s'en distingue en ce qu'elle porte sur des réplicateurs d'un type particulier : ce ne sont pas des gènes associés à leurs phénotypes mais des entités culturelles qui sont sélectionnées. Ces dernières, souvent appelées « mèmes » (pour rappeler l'analogie avec les gènes) peuvent être des gestes, des idées, des concepts, des pensées, des airs de musique, des artefacts, des normes de comportement, etc. Les entités culturelles peuvent être transmises sous forme de copies mais à la différence des gènes, cela ne se fait pas au moyen de la reproduction (transmission du matériel génétique d'un organisme porteur à un autre), mais de l'apprentissage social et en particulier de l'imitation. A l'image des gènes, les entités culturelles peuvent s'associer à d'autres pour renforcer la probabilité de leur transmission (c'est par exemple le cas des ensembles de croyances que forment les systèmes religieux). Enfin, quoiqu'analogue au mécanisme de sélection génétique, la

sélection culturelle est autonome par rapport à l'évolution génétique.

La théorie de la sélection culturelle classique a été accusée à juste titre de trop forcer l'analogie avec la sélection naturelle. Voici quelques objections auxquelles se heurte cette conception trop littérale de l'analogie.

Il est difficile de considérer les éléments culturels (ou mèmes) comme des réplicateurs au même titre que les gènes. Tout d'abord, si l'on comprend bien ce qu'est un gène et son phénotype, ce n'est pas le cas du mème ; au niveau des phénomènes culturels, on ne sait pas trop faire la différence entre le réplicateur et le phénotype. Si l'on considère une idée ou une pensée, il est à la rigueur possible de dire que c'est le phénotype d'une structure neuronale sous-jacente (MAYNARD SMITH & SZATHMARY 2000/1999) ; mais qu'en est-il des gestes, des comportements ou des artefacts ? D'autre part, même si l'on décide de focaliser l'attention uniquement sur les idées comme entités culturelles, il est très improbable qu'elles (ou plutôt leurs structures neuronales sous-jacentes) puissent être répliquées à l'identique d'un cerveau à l'autre (SPERBER 1996 ; ATRAN 2001). Le jeu bien connu du bouche à oreille suggère qu'une idée ne peut pas être transmise intacte, c'est-à-dire qu'elle n'est pas un objet dont il est possible de produire des copies exactes. Or pour que l'on puisse parler de sélection, il faut que le taux de mutation ne soit pas trop haut, c'est-à-dire que les entités culturelles restent suffisamment stables pour être présentes du début à la fin du processus de sélection ; sans une certaine stabilité au fil des réplications, rien ne peut être sélectionné. On peut donc se demander s'il y a réellement une évolution culturelle au sens darwinien.

D'autre part, l'analogie avec les gènes sous-tend l'idée d'un lien généalogique entre les différentes entités culturelles. Mais ce lien est problématique (SPERBER 1996). D'une part, il implique que pour chaque entité culturelle, il y a un parent. Or il est souvent difficile de savoir de qui l'on tient une idée (en particulier les croyances qui ont un contenu sémantique très général). D'autre part, il implique la formation d'un arbre généalogique des genres ou des espèces, qui seraient par exemple les cultures ou les langues. Or les cultures viennent sans cesse se refondre les unes dans les autres si bien qu'il n'y a pas vraiment de sens à parler d'arbre généalogique.

Une dernière critique (qui rejoint ce qui a déjà été dit plus haut) concerne la transmission des entités culturelles. A compter que l'on considère uniquement les entités qui peuvent être transmises d'un esprit à l'autre (idées, pensées, croyances), une conception de l'évolution culturelle conçue comme entièrement autonome par rapport à

l'évolution génétique (DAWKINS 1996/1976 ; BLACKMORE 1999) considère les individus comme de simples réceptacles passifs (des véhicules) des entités culturelles. Cette vision de la manière dont fonctionne la communication est extrêmement caricaturale. Dan SPERBER (1996) l'a bien montré : lorsque nous formons une idée dans notre esprit, il y a toujours une bonne part de reconstruction par rapport au modèle observé. Cette activité de reconstruction implique non seulement que l'idée modèle ne peut pas être répliquée à l'identique mais également que nous ne sommes pas des réceptacles passifs des entités culturelles ; toute idée dépend de la manière dont est constitué l'esprit qui l'a forgée. Il s'agit ici d'une critique contre l'idée même d'imitation telle qu'elle a été présentée à la section précédente ; une objection dont les théoriciens évolutionnistes devront tenir compte s'ils veulent proposer un modèle acceptable de l'évolution culturelle.

Au fond, le problème de la théorie de l'évolution culturelle comprise comme strictement analogue à l'évolution biologique tient à ce qu'elle cherche à faire des entités culturelles, des éléments complètement indépendants des supports intentionnels que sont les êtres humains.

1.2.3. La théorie de la coévolution gène-culture

Pour pallier les défauts liés à une analogie stricte entre la sélection culturelle et la sélection génétique, les théories de la « coévolution gène-culture » ou de la « double hérédité » (*dual-inheritance theory*) conçoivent les productions culturelles comme étant à la fois génétiquement et socialement transmises. Les tenants de cette approche (CAVALLI-SFORZA & FELDMAN 1981 ; BOYD & RICHERSON 1985 ; HENRICH & MCELREATH 2003) défendent l'idée que la sélection culturelle est un système d'hérédité semi-autonome, c'est-à-dire que la manière dont les entités culturelles sont transmises est largement influencée par des mécanismes psychologiques génétiquement déterminés.

Cette nouvelle approche se focalise généralement sur une sorte d'éléments culturels : les *représentations* (qui doivent être comprises au sens très large d'états mentaux aussi divers que les idées, croyances, schémas de pensée, fantaisies, désirs ou intentions). Il semblerait en effet que la culture s'effectue en priorité dans les esprits car ce sont les états mentaux des gens qui causent la production d'entités culturelles

observables comme les artefacts, les paroles, les chants, les coutumes ou les comportements. A leur tour, ces entités causeront des représentations mentales chez les individus qui les observent.

D'autre part, la nouvelle approche défend l'idée d'une analogie lâche entre la transmission culturelle et génétique. Joseph HENRICH et Robert BOYD (2002) par exemple admettent que les représentations sont rarement, voire jamais transmises à l'identique d'un esprit à l'autre. Ainsi, ils comprennent le phénomène clé de l'imitation comme une capacité de reproduire quelque chose de similaire plutôt que comme une capacité de copier à l'identique ; ils acceptent un taux non négligeable d'erreur de transmission des représentations d'une personne à une autre.[42] Mais selon eux, cette réalité n'exige pas que l'on renonce à l'idée de sélection culturelle de type darwinienne ; même si l'on ne peut pas parler de réplication de représentations d'un individu à l'autre, il est possible de retrouver une sorte de dynamique de réplicateurs au niveau de la population dans son ensemble. Voyons comment cela est possible.

Pour qu'il y ait sélection, il faut une distribution de représentations ou au moins de classes distinctes de représentations très semblables dans l'ensemble de la population. Cette stabilité sur la longue durée des entités culturelles présentes dans une population peut être garantie par l'action de biais psychologiques assez puissants. Les biais psychologiques peuvent être décrits comme des mécanismes innés, plus précisément des dispositions communes aux êtres humains qui orientent leurs choix et la manière dont ils acquièrent de l'information. La liste précise de ces biais, leur impact concret et la manière de les catégoriser font l'objet de débats (RICHERSON & BOYD 2005 ; SPERBER & CLAIDIERE 2008), mais grossièrement, on peut admettre qu'il existe deux sortes de biais (je reprends ici la distinction proposée par Dan SPERBER[43] et

[42] Certains penseurs sont réticents à l'idée que des pensées puissent être transmisses d'un esprit à l'autre via un mécanisme d'imitation (SPERBER). Je ne partage pas ces réserves. Si l'on admet une notion suffisamment large d'imitation, il me semble tout à fait possible de dire qu'un esprit peut copier le contenu d'autres esprits. Dans le cours d'une discussion avec un ami, je peux reprendre ses idées à mon compte ; en ce sens, je copie les représentations mentales de mon ami.

[43] Pour être strict, il ne faudrait sans doute pas classer Dan SPERBER (pas plus que des auteurs comme Pascal BOYER ou Scott ATRAN) dans le camp des défenseurs de la théorie de la coévolution gène-culture. Ces auteurs admettent l'idée de coévolution gène-culture mais au contraire de penseurs comme Robert BOYD, Peter RICHERSON ou Joseph HENRICH, ils sont plus réticents à l'idée que la dynamique des populations puisse réellement expliquer le phénomène de transmission et de propagation d'entités culturelles. Cela tient probablement au fait que le premier groupe de chercheurs est plus intéressé aux

CLAIDIERE 2008) : ceux qui portent sur le contenu des entités culturelles et ceux qui portent sur leur provenance.

Les « biais de contenu » (*content-based bias*) portent sur le contenu des entités culturelles ; ce sont des systèmes psychologiques qui influencent la manière dont nos esprits intègrent l'information reçue et qui favorisent par conséquent la transmission de certains types particuliers de représentations plutôt que d'autres. Une manière de comprendre ce phénomène est de recourir à la « thèse de la modularité massive »[44] selon laquelle l'architecture cognitive humaine est divisée en sous-systèmes (modules) hautement spécialisés dans le traitement de certaines informations (SPERBER 1996 ; SPERBER & HIRSCHFELD 2004). Chacun de ces modules est une réponse adaptative à certains défis posés par l'environnement dans lequel évoluaient nos ancêtres, si bien qu'ils sont relativement autonomes et spécifiques aux domaines dans le cadre desquels ils ont évolué. Parmi ces mécanismes, il y a ceux des émotions, de la reconnaissance des visages, du choix des partenaires sexuels, de l'acquisition du langage, de l'attribution d'états mentaux (théorie de l'esprit), d'une biologie naïve, etc.[45] Notons que la modularité massive n'implique pas que tout soit inné ; beaucoup de modules sont des modules d'apprentissage.

Un grand nombre de modules sont des biais de contenu. Grâce à eux, nous ne formons pas nos représentations mentales n'importe comment ; face à certains *inputs* provenant de l'environnement, les mécanismes psychologiques dirigent nos esprits vers un ensemble de représentations plutôt que vers un autre. En ce qui concerne le langage par exemple, la présence du biais de l'acquisition du langage (ce que CHOMSKY 1965 appellerait une grammaire universelle) explique pourquoi les enfants appartenant à la même communauté linguistique finissent par avoir une grammaire mentale très

---

bases cognitives de la culture qu'à sa « dynamique populationnelle ». Récemment, les deux groupes ont cependant commencé à collaborer (voir notamment HENRICH & BOYD 2002 ; CLAIDIÈRE & SPERBER 2007). Nous pouvons espérer dans les prochaines années une unification de ce domaine de recherche.

[44] Il n'est cependant pas certain que des auteurs comme Robert BOYD ou Peter RICHERSON admettraient une telle explication.

[45] La modularité de l'esprit vaut aussi bien pour les capacités (i) d'avoir des croyances simples, (ii) d'avoir des croyances sur d'autres croyances (méta-cognition), (iii) d'interpréter les croyances des autres (théorie de l'esprit), (iv) de ressentir certaines émotions spécifiques et (v) de comprendre les émotions ressenties par autrui (empathie).

similaire alors même que chacun d'entre eux aura entendu et répété un ensemble de phrases très différent.

Au fond, les biais de contenu induisent un apprentissage sélectif. Ils imposent aux représentations qui découlent de certains inputs (en ce qui nous concerne, les inputs sont les manifestations de représentations) de graviter autour d'un même espace de possibilités. Ainsi, même si pour chaque cas particulier, il y a erreur de transmission, il ne s'ensuit pas une dispersion complète de l'information contenue dans les inputs. En d'autres termes, même si chaque individu possède une variante personnelle d'une certaine représentation (entité culturelle), les biais de contenu poussent les gens à former des représentations très similaires. Et si l'on observe le phénomène à l'échelle de la population, on observera une certaine stabilité des représentations et de leur transmission.

Quant aux « biais de transmission », ils exploitent les caractéristiques des individus modèles ou les fréquences des représentations alternatives. Les deux plus importants semblent être le biais du conformisme et le biais du prestige (BOYD & RICHERSON 1985 ; HENRICH & BOYD 2002).[46] Le biais du prestige (HENRICH & GIL-WHITE 2001) est une tendance à acquérir les représentations endossées par les individus prestigieux de notre entourage.[47] Par exemple, si une star défend une cause humanitaire,

[46] Adam SMITH soulevait déjà cette réalité : « Cette disposition naturelle à accorder et à rendre semblables autant que possible nos sentiments, nos principes et nos affections, à ceux que nous voyons établis et enracinés chez les personnes avec lesquelles nous sommes obligés de vivre et de converser la plupart du temps, est la cause des effets contagieux de la bonne comme de la mauvaise compagnie. » (2003/1759, p. 311)

[47] Une explication de l'évolution de ce biais pourrait être la suivante. Les individus varient en matière de compétences, de stratégies ou de préférences. Si ces différences affectent la *fitness* des différents individus et que certaines composantes de ces différences peuvent être acquises via l'apprentissage culturel, alors la sélection naturelle peut favoriser les capacités cognitives qui poussent les individus à imiter de préférence les comportements des individus qui ont le plus de succès. Si la variation entre les compétences que l'on peut acquérir est grande et que ces compétences sont difficiles à acquérir par le biais de l'apprentissage individuel, alors il devient très intéressant d'imiter simplement les individus qui ont du succès et du prestige (déterminé en fonction d'indicateurs indirects comme la santé, le nombre de descendants ou la richesse) ; ce procédé augmente les chances d'acquérir à moindre coût des stratégies, comportements, compétences adaptés à l'environnement. Cette tendance à imiter un modèle provient simplement de l'établissement d'une connexion entre deux types de capacités préexistantes : l'imitation et

ses fans auront tendance à l'imiter. Il convient de remarquer que le biais du prestige est lié à l'épineuse question du choix des modèles ; ce dernier dépend de beaucoup de facteurs[48] et ne se fait pas toujours de manière optimale (pour des exemples concrets voir SRIPADA & STICH 2005, pp. 150-155 ; BARKOW 1989).[49] Le biais du conformisme (HENRICH & BOYD 1998) est une tendance à adopter les représentations en fonction de la fréquence de leurs occurrences (ou plutôt la fréquence des manifestations de leurs occurrences). Par exemple, si la majorité des individus du groupe dont je fais partie font des offrandes à un dieu (ce qui est une manifestation de la croyance en Dieu), j'aurai tendance à croire en ce dieu et à lui faire également des offrandes.[50] A l'aide de modèles mathématiques, RICHERSON et BOYD (1985) ont montré que les biais de conformisme et de prestige peuvent, au niveau de la population, compenser les insuffisances des processus de préservation de l'information au niveau individuel ; une entité culturelle imitée à grande échelle restera stable au niveau de la population même si au niveau individuel, il y a des erreurs de transmission.

La stabilité au niveau populationnel des entités culturelles étant garantie par les biais psychologiques, il est désormais possible d'imaginer la possibilité d'une dynamique de sélection culturelle partiellement indépendante de la sélection biologique ; cette dernière influence l'évolution culturelle au moyen des biais psychologiques. A la section 2.3, nous verrons des exemples concrets de ce type de

---

la capacité d'établir une hiérarchie entre différents membres d'un groupe. Notons que cette dernière est déjà présente dans un grand nombre d'espèces.

[48] Il existe des études psychologiques sur cette question. Jody DAVIS et Caryl RUSBULT (2001) par exemple ont travaillé sur le phénomène d'alignement d'attitude (*attitude alignment*), une tendance à calquer nos comportements sur ceux de nos proches et de nos amis. Jonathan HAIDT (2001) mentionne différentes théories et expériences qui éclairent la manière dont les êtres humains sont influencés par leurs pairs (voir aussi HENRICH & BOYD 1998 ; RICHERSON & BOYD 1985).

[49] Au cinquième siècle avant J.-C. déjà, le présocratique XÉNOPHANE s'indignait de l'influence exercée par les sportifs sur la population, alors qu'elle aurait dû prendre exemple sur les intellectuels dont il se considérait comme un brillant exemple (B I ; Fragment II, pp. 114-115).

[50] Une explication de l'évolution de ce biais pourrait être la suivante : dans un environnement pauvre en information (où il n'est pas évident de juger du succès effectif des individus si bien qu'il est difficile d'en choisir certains comme modèles) et où l'apprentissage individuel est coûteux, il vaut la peine de se conformer simplement à la majorité car les comportements de la majorité contiennent implicitement les effets de chaque expérience et effort d'apprentissage individuel.

processus (où au niveau populationnel, il y a sélection d'entités culturelles au détriment d'autres entités).

La théorie de la coévolution gène-culture tire son nom du fait qu'elle considère non seulement l'influence de l'évolution biologique sur l'évolution culturelle mais également l'influence inverse : si un trait (par exemple un comportement, la maîtrise d'une nouvelle technique, etc.) est acquis par l'apprentissage ou l'imitation et qu'il est adapté à un environnement qui reste stable relativement à cette adaptation, alors il se peut que sur le long terme cela exerce un impact sur l'évolution de certains traits génétiquement codés (BARKOW 1989). C'est ce que l'on appelle souvent « l'effet Baldwin » en référence à l'homme qui a découvert ce phénomène (1896). Plus récemment, la notion de « construction de niche » (*niche construction*) a également été utilisée pour se référer à ce phénomène (DEACON 1997; WEBER & BRUCE 2003 ; LACHAPELLE *et al.* 2006).[51] La construction de niche met en valeur l'idée que les organismes peuvent modeler à leur avantage l'environnement dans lequel ils vivent, et par là modifient les pressions sélectives qui agissent sur ces mêmes organismes ; sur le long terme il peut en résulter des changements au niveau génétique. Par exemple, au cours de l'histoire humaine les ancêtres des contrées occidentales ont commencé à tenir du bétail et à en consommer le lait. A ce stade d'évolution, leur constitution génétique ne leur permettait pas de digérer ces produits correctement ; mais en persévérant dans cette habitude de consommation, au fil des générations, une tolérance au lactose a été sélectionnée. Cela explique pourquoi les populations asiatiques qui n'ont pas connu la même histoire sont moins tolérantes aux produits laitiers (DEACON 1997). Nous verrons à la section 3.4 que des contextes sociaux coopératifs issus d'une pression culturelle peuvent causer la fixation de gènes responsables des traits altruistes psychologiques.

Conclusion

Dans ce chapitre, nous avons vu que la théorie de l'évolution est un modèle explicatif qui ne s'applique pas uniquement au monde biologique, mais également au monde culturel. Il est donc important de ne pas confondre « explication évolutionnaire »

[51] Pour une discussion sur les rapports entre la *construction de niche* et la notion (au contenu très similaire) de *phénotype étendu* développée par DAWKINS (1999/1982), voir LALAND 2004 et DAWKINS 2004.

et « explication biologique basée sur les gènes » ; toute explication évolutionnaire ne tient pas forcément compte des gènes et de leurs phénotypes.

D'autre part, j'ai tenté d'éclairer les liens entre le monde culturel et l'évolution biologique. Trois liens ressortent de cette analyse. Tout d'abord, la genèse de la culture peut être expliquée en termes de sélection naturelle et d'adaptation génétique. En effet, l'existence de la culture nécessite certaines pré-conditions ; en l'occurrence, la possession, par les « êtres de culture », de facultés cognitives indispensables à la transmission d'entités culturelles. Parmi ces facultés, la plus importante semble celle de l'imitation (qui ne doit cependant pas être comprise comme une capacité de copier fidèlement les entités culturelles mais plutôt à en reconstruire de similaires). Ces mécanismes cognitifs sous-jacents à la culture sont le résultat d'un processus naturel d'évolution. Deuxièmement, il est possible de parler d'évolution culturelle (au sens où il existe des propriétés culturelles héritables qui évoluent en un sens darwinien) en tant que processus semi-autonome par rapport à l'évolution biologique. L'évolution culturelle est influencée par un bon nombre de biais psychologiques génétiquement déterminés ; et ces biais sont garants de la dynamique de sélection culturelle. Enfin, l'évolution culturelle, par effet Baldwin, peut elle-même exercer un impact au niveau de l'évolution biologique.

2. L'altruisme évolutionnaire

Les bases de la théorie de l'évolution étant posées, nous pouvons faire un premier pas en direction de l'éthique en analysant la manière dont les biologistes abordent les comportements altruistes.

Dans ce chapitre, nous verrons que l'altruisme pose un problème aux biologistes car il semble remettre en question la théorie de l'évolution elle-même ; en effet, comment peut-on expliquer que des comportements par définition défavorables à l'individu qui les développe puissent être sélectionnés ? Ce problème a occupé les biologistes de l'évolution depuis Darwin. Un certain nombre d'explications (pour la plupart compatibles entre elles) ont été proposées pour rendre compte de l'évolution de l'altruisme dans le monde animal et humain ; les plus importantes seront présentées et discutées dans ce chapitre. Toute la discussion se mènera sur un fond de controverse historique entre les penseurs qui soutiennent l'existence de l'altruisme dans le monde biologique et ceux qui pensent qu'à y regarder de plus près, cet altruisme n'est qu'apparent. En fin de compte, nous verrons que d'un point de vue théorique, cette controverse n'est que du vent.

2.1. Paradoxe et controverse autour de l'altruisme évolutionnaire

Dans les écrits évolutionnaires, l'altruisme est défini en termes de *fitness*. En voici une première définition.

Un comportement est dit altruiste s'il a pour effet d'augmenter la fitness d'autrui aux dépens de la fitness de l'individu qui développe ce comportement.[52]

Quelques éclaircissements terminologiques s'imposent ici. Tout d'abord, dans ce contexte évolutionnaire, *autrui* doit être compris comme un ou plusieurs individus de la même espèce que l'agent altruiste.

[52] Nous verrons que cette définition est suffisamment floue pour être comprise de plusieurs manières.

De plus, lorsque l'on parle d'un *individu altruiste*, on pense à un individu dont la composition génétique induit une tendance plus ou moins forte à adopter un comportement altruiste.

D'autre part, la notion d'altruisme évolutionnaire n'a de sens que dans un contexte bien défini, où l'on compare le résultat de différents comportements sur la *fitness* des individus concernés ; dans ce contexte les individus qui agissent au détriment de leur propre *fitness* sont considérés comme altruistes par rapport à ceux qui n'agissent pas de la sorte (ces derniers étant parfois qualifiés d'égoïstes). Ainsi, un comportement ne peut être considéré comme altruiste qu'en comparaison avec d'autres comportements pratiqués dans une population.

Un autre point important à souligner est le fait que l'altruisme évolutionnaire se calcule uniquement en fonction des *conséquences* des comportements sur la *fitness* des organismes ; cette notion est donc très éloignée de celle qui est généralement utilisée par les philosophes et les psychologues ; ces derniers s'intéressent davantage aux *motivations* qui ont poussé les agents à agir de manière altruiste (à ce propos, voir les sections 3.1 et 3.2).

Enfin, on trouve souvent dans la littérature, le terme de « gène pour l'altruisme » (*gene for*) ou « gène qui code pour l'altruisme » ; cette manière de s'exprimer est un raccourci qui ne signifie pas ce qu'il a l'air de signifier. En effet, les biologistes et généticiens s'accordent sur le fait que ce n'est pas un seul gène qui induit un comportement altruiste ; un comportement altruiste est un phénomène extrêmement complexe dans lequel est impliquée une portion plus ou moins grande de matériel génétique, diversement répartie sur les séquences d'ADN (sans oublier toutes les complications liées à la transcription de ce matériel génétique). Par rapport à ce comportement, le « gène pour » désigne « ce qui fait que » (ou la portion génétique qui fait que) un individu se comporte de manière altruiste plutôt que de manière non altruiste.[53] Il est tout à fait concevable que ce « ce qui fait que » consiste en une différence minime (mais dont les conséquences peuvent être grandes !) qui se greffe sur un comportement préexistant ; par exemple un comportement de soin parental de type « je nourris mes propres petits » qui se transforme, par la magie d'une petite mutation

[53] Notons également que ce n'est pas parce qu'un individu possède un code génétique qui induit un comportement altruiste qu'il développera forcément ce comportement dans la réalité ; les contingences environnementales et développementales peuvent empêcher la réalisation du code.

génétique d'un allèle, en soin généralisé de type « je nourris tous les petits que je rencontre ». Dans les mots de DAWKINS :

> « Un comportement altruiste peut s'avérer très complexe, mais cette complexité n'est pas due à un nouveau gène mutant, mais à des processus développementaux préexistants sur lesquels agit ce gène. Avant l'arrivée du nouveau gène, un comportement complexe existe déjà ; il est le résultat d'un processus développemental long et complexe qui implique de nombreux gènes ainsi que des facteurs environnementaux. Le nouveau gène lui fait simplement prendre un nouveau virage qui aura des effets phénotypiques cruciaux. Par exemple, ce qui était un comportement de soin maternel [*maternal care*] complexe est devenu un comportement complexe de soin dirigé vers n'importe quel petit [*sibling care*]. La transition d'un soin prodigué de manière discriminatoire envers ses propres petits à un soin non discriminatoire est une transition simple, même si les comportements sur lesquels elle opère sont très complexes. » (DAWKINS 1979, p. 190, ma traduction)[54]

Après ces éclaircissements, reprenons la définition du comportement altruiste et considérons un exemple : Selon la définition de la *fitness* classique donnée à la section 1.1.1 (p. 22), la *fitness* d'un individu se calcule en termes de viabilité (capacité de survivre) et de fécondité (capacité de produire une descendance). Imaginons maintenant une abeille qui possède un code génétique qui lui dicte le comportement suivant : « Si tu vois un intrus qui s'approche du nid, lance-toi sur lui et pique-le ! » Sachant que l'abeille meurt après avoir piqué, on peut considérer qu'il s'agit d'un individu altruiste (dans le sens où cet individu adopte un comportement altruiste) ; en effet, si l'abeille pique, elle voit sa viabilité réduite à néant tout en augmentant les chances de survie de ses congénères.

[54] « Altruistic behaviour may be very complex, but it got its complexity, not from a new mutant gene, but from the pre-existing developmental process that the gene acted upon. There already was complex behaviour before the new gene came along, and that complex behaviour was the result of a long and intricate developmental process involving a large number of genes and environmental factors. The new gene of interest simply gave this existing complex process a crude kick, the end result of which was a crucial change in the complex phenotypic effect. What had been complex maternal care, say, became complex sibling care. The shift from maternal to sibling care was a simple one, even if both maternal and sibling care are very complex in themselves. » (DAWKINS 1979, p. 190)

Les comportements altruistes ont beaucoup intéressé les biologistes car ils mènent au fameux paradoxe de l'altruisme. Considérons une population composée d'individus altruistes et d'individus non altruistes. Cette population compte un nombre relativement restreint d'individus évoluant dans le même milieu. Imaginons que ces individus soient semblables du point de vue de leur adaptation à l'environnement et de leurs capacités reproductrices, à une différence près : de temps en temps, les altruistes adoptent un comportement qui a pour effet de favoriser les autres individus du groupe au détriment de leur propre *fitness* (par exemple en distribuant sans discrimination une grande partie de leur nourriture de base). En agissant de la sorte, les altruistes perdent un peu de viabilité et de fécondité potentielles tout en permettant à un bon nombre de leurs congénères de se développer et d'investir de l'énergie dans leur propre descendance.

Il s'ensuit que, si à la fin de la première génération, on compare les *fitness* respectives (au sens de *fitness* classique) d'un individu altruiste et d'un individu non altruiste, on remarque que, quelle que soit la *fitness* de l'altruiste, elle sera inférieure à celle du non-altruiste. En d'autres termes, cela signifie que le nombre moyen de petits par individu non altruiste sera supérieur au nombre moyen de petits par individu altruiste. Et cela reste valable quelle que soit la proportion d'altruistes au sein du groupe.[55] Ainsi, à la génération suivante, la proportion de non-altruistes aura augmenté par rapport à celle des altruistes, et ainsi de suite de génération en génération jusqu'à la disparition complète des altruistes.

A ce propos, il faut remarquer que plus il y a d'altruistes dans le groupe, plus la *fitness* moyenne des individus de ce groupe sera haute (à condition que les actions altruistes génèrent un bénéfice global en termes de *fitness* qui soit supérieur à la perte de *fitness* individuelle). Et inversement, si une population donnée passe d'une majorité d'altruistes à une majorité de non-altruistes (ce qui est le cas dans notre exemple), alors la *fitness* moyenne des individus de cette population baisse. Voilà une raison supplémentaire de s'attrister de la disparition des altruistes !

Le modèle ci-dessus démontre bien que, quelle que soit la population considérée, le trait de l'altruisme est tragiquement voué à l'extinction. Ce phénomène est dû à la pression de la sélection naturelle au fil des générations. En d'autres termes, la théorie de

[55] En réalité, plus il y a d'altruistes dans une population, mieux se portent les égoïstes (puisqu'ils profitent des avantages sans rien donner en retour).

l'évolution semble prédire que les comportements altruistes, par définition, ne peuvent pas supporter la pression de la sélection naturelle ; ils sont forcément voués à la disparition au profit des comportements non altruistes (et cela même si du point de vue du groupe, il est préférable que des individus soient altruistes). Or, et c'est ici qu'apparaît le paradoxe, on peut observer dans le monde animal des comportements altruistes, ou du moins qui ont tout l'air d'être altruistes ! L'exemple des abeilles kamikazes cité plus haut n'est pas un hasard ; il s'agit d'une observation empirique devant laquelle DARWIN lui-même est resté muet. Comme autre exemple rebattu (dont tout randonneur aura certainement fait l'expérience), on peut citer celui des marmottes siffleuses : dans chaque groupe de marmotte il y a des sentinelles qui mettent leur vie en danger en sifflant pour avertir leurs congénères de l'approche d'un prédateur (le sifflement attire l'attention du prédateur). Ainsi, les théoriciens de l'évolution se trouvent devant un problème dérangeant ; leur fameuse théorie ne permet pas d'expliquer un type de phénomènes empiriquement observable.

Depuis la découverte, par DARWIN, du paradoxe de l'altruisme, une légion de penseurs a tenté de le résoudre. C'est l'histoire de ces tentatives qui sera présentée dans cette section. Globalement, deux grandes tendances se profilent.

D'un côté, on trouve ceux que je nommerai les « stratèges de l'égoïsme », qui affirment que tous les comportements altruistes évolutionnaires (comme ceux des abeilles kamikazes ou des marmottes siffleuses par exemple) peuvent être compris en termes d'avantages sélectifs. Nous verrons que les défenseurs de cette position sont ceux qui adoptent la perspective du gène (HAMILTON, DAWKINS, R. FISHER, TRIVERS, E. WILSON, etc.).

De l'autre côté, on trouve ceux que je nommerai les « romantiques » parce qu'ils persévèrent dans la pensée que l'altruisme évolutionnaire est effectivement sélectivement désavantageux. Nous verrons que les partisans de cette approche s'appuient sur des théories de sélection de groupe afin de montrer comment les comportements altruistes, par définition défavorables du point de vue de la *fitness* des agents, peuvent passer le cap de la sélection naturelle (LORENZ, D. WILSON, etc.). Jusqu'à ce jour, les partisans de ces deux approches s'affrontent à coups d'arguments.

Je qualifierai cette bataille de « controverse autour de l'altruisme évolutionnaire »[56] et tâcherai de montrer qu'en fin de compte, cette controverse n'est que du vent : si l'on prend en compte les différents niveaux de la sélection tout en s'accordant sur la signification exacte du terme « altruisme évolutionnaire », elle disparaît d'elle-même.

2.2. L'altruisme évolutionnaire dans le monde animal

2.2.1. La sélection de parentèle

Le paradoxe de l'altruisme semble montrer que les comportements altruistes ne peuvent en principe pas être sélectionnés au fil de l'évolution. C'est du moins la conclusion à laquelle on est réduit si l'on cherche à comprendre le mécanisme de la sélection naturelle en termes de *fitness* classique. Darwin, le premier à prendre conscience de ce paradoxe, a cherché la solution dans une théorie de la sélection de groupe. Nous reviendrons plus tard sur cette approche. Mais procédons à un saut chronologique et commençons par relater la solution proposée par William Hamilton dans les années 1960 (HAMILTON 1964).

i. Le point de vue du gène et la fitness inclusive

La biologie de l'évolution des années 1960 se caractérise par l'adoption d'une nouvelle perspective : celle du gène. Désormais, on admet que les traits héréditaires ne se transmettent pas eux-mêmes, mais par l'intermédiaire de gènes. Ainsi, il faut toujours garder en tête la relation gène / phénotype (ce dernier étant une manifestation observable d'un ou plusieurs gènes).

Prendre la perspective du gène signifie penser que la sélection naturelle favorise les gènes capables de se répliquer (créer des copies exactes d'eux-mêmes) plus que les autres. Ainsi, lorsqu'on cherche à comprendre les raisons pour lesquelles un phénotype a été sélectionné, on doit tenter de comprendre quels étaient les avantages sélectifs des gènes qui induisent ce phénotype. La viabilité et la fécondité des individus porteurs de

[56] Précisons que cette controverse porte sur la possibilité de la sélection de *types de comportements authentiquement altruistes* (au sens évolutionnaire du terme). Elle ne porte pas sur la possibilité de l'existence d'*actions* altruistes évolutionnaires isolées ; personne ne nierait leur occurrence occasionnelle.

ces gènes ne sont prises en considération que de façon indirecte, dans la mesure où elles favorisent la réplication des gènes considérés. Une conséquence importante de ce nouveau point de vue, est qu'en accordant une priorité à « l'intérêt » des gènes (on se demande comment un gène qui induit un phénotype peut se répandre dans l'ensemble du pool génétique d'une population[57]) par rapport à l'intérêt des individus porteurs des gènes, on est conduit à rejeter la notion traditionnelle de *fitness* classique (p. 22). En effet, la *fitness* classique est une manière de calculer les avantages sélectifs en tenant compte du succès reproductif des individus (c'est-à-dire la capacité d'un individu à transmettre ses propres traits à ses enfants, petits enfants, etc.) et non du succès de réplication des gènes. Dès lors, il s'agit de réviser cette notion de *fitness* et de l'interpréter de manière à rendre compte du fait que la sélection naturelle opère sur les gènes et leurs effets phénotypiques, abstraction faite des individus porteurs de ces gènes. C'est ce qu'a fait William HAMILTON en définissant ce qui a été baptisé, la *fitness* inclusive.[58]

La *fitness* inclusive est une mesure qui cumule la viabilité et la fécondité individuelle, et les effets du comportement de l'individu focal sur la viabilité et la fécondité de ses proches parents, chaque parent comptant en proportion du coefficient d'apparentement. Ce calcul, plus complexe que celui de la *fitness* classique, comporte deux étapes : premièrement le calcul de l'apparentement génétique, ensuite l'application de la règle de Hamilton. Considérons ces deux étapes :

HAMILTON a établi une manière de calculer le coefficient d'apparentement[59] entre deux individus. On assimile souvent ce coefficient au simple degré de parenté par voie de descendance.[60] Ce dernier correspond à la proportion moyenne du génome partagé

[57] Le pool génétique d'une espèce ou d'une population est l'ensemble des gènes (plus particulièrement des allèles) véhiculés par les individus vivants de cette espèce ou de cette population.

[58] HAMILTON (1963) mentionne explicitement qu'il puise les lignes principales de sa théorie chez HALDANE (1932).

[59] Pour désigner le coefficient d'apparentement de HAMILTON, bien des termes sont utilisés : *coefficient of relationship* (HAMILTON 1963 *;* DAWKINS 1999/1982, p. 189*), coefficient of relatedness* (DAWKINS 1979, p. 191 ; PEPPER 2000), *degree of relatedness* (J. KREBS & DAVIES 1993/1981, p. 265 ; RIDLEY 2004/2003, p. 242), *relatedness* (GRAFEN 1985). En français on trouve les termes de *degré de parenté* (DAWKINS 1996/1976, p. 131) et *coefficient de parenté* (PERRIN 2005, p. 54).

[60] Il s'agit en réalité du calcul de MALECOT (1948) qui, comme on le verra plus loin, ne devrait pas être confondu avec le coefficient d'apparentement. Notons que MALECOT fonde son coefficient sur la notion

par deux individus, par voie de descendance. Et si l'on adopte une perspective géno-centrée, il correspond à la probabilité que deux individus possèdent une copie d'un même gène, par voie de descendance.[61] Par exemple, dans la plupart des espèces animales, le degré de parenté par voie de descendance entre un parent et son enfant est de 50% (pour chacun des gènes du parent, il y a 50% de chances qu'il ait été transmis à l'enfant).

Richard DAWKINS (1996/1976, pp. 127-152), John KREBS et Nicholas DAVIES (1993/1981, pp. 265-269) présentent de manière didactique la règle générale pour déterminer le degré de parenté par voie de descendance entre deux individus A et B (à ce propos, voir le schéma ci-dessous). En se représentant un arbre généalogique, il faut d'abord identifier tous les ancêtres communs à A et B ; par exemple, les ancêtres communs de deux cousins sont la grand-mère et le grand-père. Ensuite il faut compter la distance entre générations, c'est-à-dire que partant de A, il faut remonter l'arbre de la famille jusqu'à ce que l'on parvienne à un ancêtre commun, puis redescendre l'arbre pour aboutir à B ; par exemple, la distance entre deux cousins est de quatre. Ensuite, on peut calculer le degré de parenté dont l'ancêtre commun est responsable en multipliant par ½ à chaque étape de la distance entre générations ; ainsi, si la distance entre générations via un ancêtre particulier est égale à *d* étapes, la portion de parenté due à cet ancêtre est de $(½)^d$; par exemple, si la distance entre générations est de quatre, le calcul est de ½ x ½ x ½ x ½ soit $(½)^4$. Si A et B ont plus qu'un ancêtre commun, il faut renouveler le calcul pour chaque ancêtre commun ; par exemple, le degré de parenté entre deux cousins est de 2 x $(½)^4$ = 1/8.[62] Pour des parentés aussi éloignées que celles des cousins de troisième degré (2 x $(½)^8$ = 1/128) nous nous rapprochons de la probabilité qu'un gène possédé par A soit partagé par n'importe quel individu pris au hasard dans la population.

---

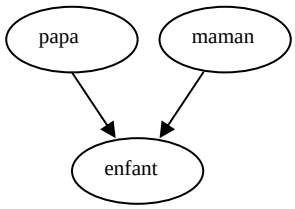
d'« identité par *ascendance* ». Je suivrai cependant la terminologie de J. KREBS, DAVIES (1993/1981) et DAWKINS (1996/1976) en parlant de coefficient de parenté par voie de *descendance*.

[61] « The probability that a parent and an offspring will share a copy of a particular gene identical by descent (in an outbreeding species) is 0.5. This quantity is called the coefficient of relatedness, often denoted by *r*. » (J. KREBS & DAVIES 1993/*1981,* p. 265)

[62] Ainsi, si les individus sont issus d'une longue série de rapports incestueux, le calcul devient très complexe puisqu'il faut connaître et calculer tous les liens de parenté.

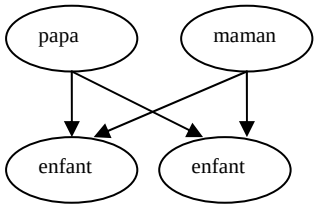
Modèle parent / enfant

1 x $(½)^1$ = ½ (0.5)



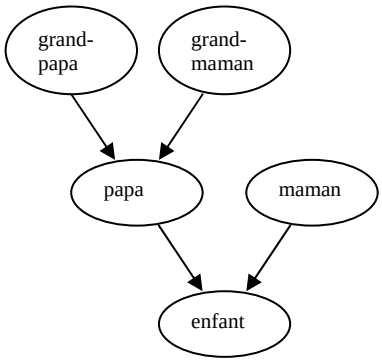
Modèle frère / sœur

2 x $(½)^2$ = ½ (0.5)



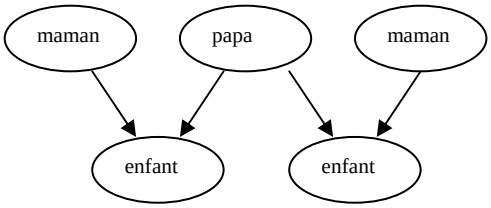
Modèle grand-parent / enfant

1 x $(½)^2$ = ¼ (0.25)



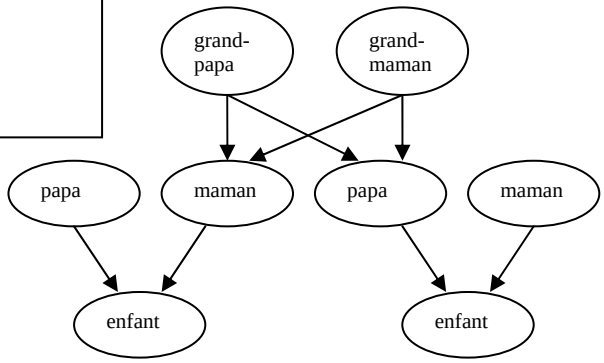
Modèle demi-frères

1 x $(½)^2$ = 1/4 (0.25)



Modèle cousins

2 x $(½)^4$ = 1/8 (0.125)



Ce schéma est inspiré de J. KREBS & DAVIES 1993/1981, p. 267

(les flèches indiquent une transmission de 50% des gènes)

Ainsi, selon cette approche, le coefficient d'apparentement correspondrait à la proportion de gènes identiques entre deux individus, due à une descendance récente calculable à l'aide d'un arbre généalogique.

Venons-en à la règle de Hamilton : il s'agit de la règle

$$rb > c$$

où

r = le coefficient d'apparentement entre le donneur et le/les receveur/s

b = le bénéfice (en termes de viabilité et fécondité) pour le/les receveur/s, qui découle de l'action altruiste

c = le coût (en termes de viabilité et fécondité) pour le donneur, qui découle de l'action altruiste

Le bénéfice (du/des receveur/s) multiplié par le coefficient d'apparentement doit être supérieur au coût (pour le donneur). En d'autres termes, cette règle signifie simplement qu'un comportement coopératif évolue si, par ce biais, les gènes responsables de ce comportement permettent au total un plus grand nombre de copies.

Illustrons cette règle à l'aide de quelques exemples simples. Posons comme conditions de départ que i/ les individus bénéficiaires de l'action altruiste sont jeunes et bien-portants et ii/ le gène responsable du comportement altruiste est rare (la raison de cette condition deviendra plus claire par la suite).

Première situation : Un individu donne sa vie pour sauver celle de trois de ses frères. Les trois frères seraient morts sans ce sacrifice. Si on applique le calcul de Hamilton à cette situation, cela nous donne les valeurs suivantes :

- Coût pour le frère altruiste = 100 (il perd la vie)
- Gain pour les trois frères sauvés = 300 (trois vies sauvées)
- Coefficient d'apparentement = 0.5 (chacun des frères sauvés partage 50% de ses gènes avec le frère altruiste).

On voit que l'inégalité est satisfaite : 0.5 x 300 > 100. Il s'ensuit qu'un comportement sacrificiel de ce type a des chances d'être sélectionné au fil de l'évolution car il permet d'augmenter les chances de réplication des gènes responsables de ce comportement. Dans le cas présent, si le premier frère ne s'était pas sacrifié, les trois autres seraient

morts et la proportion de copies du gène qui pousse au sacrifice, aurait baissé dans le pool génétique.

Seconde situation : Un individu donne sa vie pour sauver celle de quatre de ses frères. Il y a une probabilité de 50% pour que chaque frère échappe au danger sans ce sacrifice. Si on applique le calcul de Hamilton à cette situation, cela nous donne les valeurs suivantes :

- Coût pour le frère altruiste = 100 (il perd la vie)
- Gain pour les quatre frères sauvés = 200 (sauvetage de quatre vies qui avaient 50% de chances de s'en sortir)
- Coefficient d'apparentement = 0.5 (chacun des frères sauvés partage 50% de ses gènes avec le frère altruiste)

On voit que l'inégalité n'est pas satisfaite. Il n'est pas le cas que 0.5 x 200 > 100. Il s'ensuit que du point de vue des gènes, ce sacrifice n'est pas vraiment intéressant. Il y a donc peu de chance pour qu'il soit sélectionné au fil de l'évolution.

Troisième situation : Un individu donne sa vie pour sauver celle d'un de ses frères. Le frère serait mort sans ce sacrifice. Si on applique le calcul de Hamilton à cette situation, cela nous donne les valeurs suivantes :

- Coût pour le frère altruiste = 100 (il perd la vie)
- Gain pour le frère sauvé = 100 (sauvetage d'une vie)
- Coefficient d'apparentement = 0.5 (les deux frères partagent 50% de leurs gènes)

On voit que la règle de Hamilton n'est pas satisfaite car 0.5 x 100 < 100. Il s'ensuit que du point de vue des gènes, ce sacrifice n'est pas intéressant car il ne pourra pas passer le cap de la sélection naturelle.

De manière très schématique, la règle de Hamilton prédit qu'un comportement altruiste envers des frères et sœurs sera sélectionné uniquement s'il occasionne un gain plus de deux fois supérieur à la perte occasionnée à l'agent ; en faveur de demi-frères, le gain devra encore doubler, etc. Ce processus est couramment appelé « sélection de parentèle ». HAMILTON le décrit en ces termes :

> « Ainsi, un gène qui cause un comportement altruiste envers des frères et sœurs sera sélectionné à la seule condition que ce comportement et les circonstances soient généralement telles que le gain s'avère plus de deux fois supérieur au coût ; pour des

> demi-frères, il doit être plus de quatre fois supérieur au coût, etc. » (HAMILTON 1963, p. 355)[63]

Précisons toutefois que si les trois exemples qui viennent d'être donnés sont séduisants par leur simplicité et permettent de saisir le fonctionnement de la règle de Hamilton, on ne peut pas leur accorder de réel crédit au niveau empirique. En effet, la sélection naturelle opère sur des types de comportements. Or il est difficile d'imaginer que des comportements sacrificiels aussi particuliers (donner sa vie pour sauver son frère en bonne santé) puissent être sélectionnés au fil de l'évolution.[64]

Sur la base de ces explications élégantes mais excessivement simplificatrices de la *fitness* inclusive, prenons le temps de considérer les détails du calcul et de la règle de Hamilton. Il est très troublant de constater que plus on s'intéresse aux finesses de la théorie de HAMILTON, plus les difficultés de compréhension s'accumulent. En effet, si l'on reconsidère le genre d'exemples proposés ci-dessus, un certain nombre de problèmes surgissent.

Il y a tout d'abord le problème du calcul des effets des actions d'un individu sur les chances de reproduction d'un autre. Si l'on veut être rigoureux le calcul de la *fitness* inclusive devrait prendre en compte uniquement des individus capables de se reproduire ; en effet, du point de vue de la propagation des gènes, une vieille mère (degré de parenté par voie de descendance : 50%) ne compte pas la même chose qu'un jeune frère (degré de parenté par voie de descendance : 50%) qui vient d'atteindre l'âge de reproduction…

Ensuite, il y a le problème du degré de rareté du gène considéré. Citons DAWKINS à cette occasion :

> « Le problème de la mesure du degré de parenté [à comprendre au sens de « coefficient d'apparentement »] fait trébucher bon nombre d'entre nous sur le point suivant. Quels que soient deux membres d'une espèce, qu'ils appartiennent ou non à la même famille, ils ont souvent en commun plus de 90% de leurs gènes. De quoi parlons-nous donc

[63] « Thus a gene causing altruistic behavior towards brothers and sisters will be selected only if the behavior and the circumstances are generally such that the gain is more than twice the loss ; for half-brothers it must be more than four times the loss ; and so on. » (HAMILTON 1963, p. 355)

[64] Pour une critique de l'usage d'exemples aussi caricaturaux, voir GRAFEN 1985, p.70-71.

> lorsque nous disons que le degré de parenté entre deux frères est de ½ ou qu'il est de 1/8 entre des cousins germains ? La réponse est que les frères partagent la moitié de leurs gènes en plus des 90% (tout ce que vous voulez) que tous les individus ont en commun de toute façon. » (DAWKINS 1996/1976, p. 382)

Le problème mentionné par DAWKINS est le suivant : Deux frères peuvent avoir 95% de leurs gènes en commun ; ces 95% sont, par exemple, composés de 90% communs à tous les membres de leur espèce et de 5% moins répandus, hérités de leurs ancêtres communs par voie de descendance. Toutefois, même s'ils ont 95% de gènes en commun, les deux frères ne partagent pas plus de 50% des gènes par voie de descendance (chacun des deux parents ayant 25% de chances de transmettre une réplique de leur propre copie de gène aux deux frères à la fois) ; et parmi ces 50%, il y a des gènes communs à tous les individus de l'espèce et il y a une petite proportion de gènes moins communs, voire rares. Dans cet ensemble de proportions de gènes héritées, DAWKINS semble nous dire que le coefficient d'apparentement entre les deux frères correspond aux 50% de chances qu'un gène, qui ne fait pas partie des 90% (ou ce que vous voudrez) que tous les individus possèdent de toute manière en commun, ait été transmis aux deux frères par voie de descendance. Voilà qui complique passablement notre affaire !

En réalité, le problème vient de ce que tout ce qui a été dit jusqu'à maintenant au sujet du calcul du coefficient d'apparentement s'appuie sur une hypothèse extrêmement lourde : on postule que les individus évoluent dans une population illimitée et ont la même probabilité d'être confrontés à n'importe quel autre individu de cette population. Dans ce contexte, le coefficient d'apparentement est considéré comme une mesure de parenté absolue. Or cette hypothèse ne correspond pas à la réalité. Afin de rendre compte de celle-ci, il faudrait tenir compte du contexte dans lequel évoluent les individus considérés ; il faudrait imaginer différents groupes d'individus plus ou moins solidaires, dont certains entrent en compétition et d'autres ne se rencontrent jamais. Ainsi, *r* devient une mesure *relative* et non absolue.[65]

[65] L'explication proposée ci-dessus par DAWKINS est peu éclairante dans la mesure où elle met en jeu des *valeurs fixes* (comme les 90% de gènes possédés par tous les individus considérés) alors qu'en réalité, les valeurs qui sont utilisées dans le calcul de *r* doivent être comprises comme des proportions moyennes ou des probabilités.

Je vais tâcher, sans entrer dans les formules mathématiques complexes, de rendre compte de la subtilité de *r* à l'aide d'un modèle de groupes[66] : *r* est une notion relative qui établit une relation entre deux individus issus d'un même voisinage social, relativement à une population de référence.[67] Le voisinage social est une notion souple qui peut être comprise comme un groupe d'individus qui interagissent régulièrement ou plus simplement comme un nombre restreint d'individus parents (dans l'exemple de DAWKINS, le voisinage social est composé de deux frères).[68] La population de référence correspond au domaine de compétition, c'est-à-dire à un ensemble d'individus de la même espèce, susceptibles d'entrer en compétition. *r* correspond à la proportion attendue de gènes communs entre un individu X et n'importe quel individu Y issu du même voisinage social[69], en excédent de la proportion moyenne de gènes communs entre X et un individu quelconque de la population de référence. Et si l'on adopte une perspective géno-centrique, *r* correspond à la probabilité qu'un gène d'un individu X se trouve également chez n'importe quel individu Y issu du même voisinage social, en excédent de la probabilité qu'il se trouve chez un individu quelconque de la population de référence.[70]

Nous pouvons tirer quatre corollaires de cette explication de *r* à travers un modèle de groupe.

Premièrement, l'impact de la sélection de parentèle est dépendant de la proportion de

[66] Dans GRAFEN 1985, on trouve une présentation géométrique du coefficient d'apparentement. Je ne la présenterai cependant pas ici.

[67] Cette explication m'a été donnée par Nicolas PERRIN (pour les détails mathématiques, voir PERRIN 2005). Pour un compte rendu plus global du phénomène, voir QUELLER 1994.

[68] Pour être plus précis, on parle souvent dans ce contexte de *trait group*, qui est un ensemble temporaire d'individus qui interagissent dans un sens écologiquement significatif ; plus précisément, un *trait group* existe tant que la *fitness* des individus qui le composent est influencée par un certain trait (typiquement, le comportement altruiste d'un agent de ce groupe). Pour plus de précisions quant au *trait group* et à la manière de calculer *r* en fonction de cette notion, voir PEPPER 2000.

[69] En général, si le voisinage social est constitué uniquement des deux individus, cette proportion correspond au degré de parenté par voie de descendance. Et si le voisinage social est plus complexe, il s'agit de la moyenne des différents degrés de parenté par voie de descendance entre les individus du voisinage en question.

[70] Nicolas PERRIN explique que le coefficient d'apparentement correspond à « la proportion de gènes en commun en excédent de ce qui est attendu par hasard dans la population de référence » (PERRIN 2005, p. 73).

gènes communément partagés dans la population de référence. Illustrons ce phénomène par un exemple. Un groupe de chiots de prairie est composé d'individus tous issus des deux mêmes parents ; les membres de ce voisinage social, par voie de descendance, entretiennent des liens de parenté plus étroits qu'avec les autres chiens de la population de référence. Imaginons maintenant deux mondes possibles dans lesquels peuvent évoluer ces chiens. Dans le premier monde, Max, un chiot du groupe de voisinage A a une probabilité moyenne de partager 80% de ses gènes avec n'importe quel chien issu de la population de référence. D'autre part, Max a un frère, Félix, avec lequel il partage 50% de ses gènes par voie de descendance. Dans le deuxième monde, Max', un chiot du groupe de voisinage A' a une probabilité moyenne de partager 95% de ses gènes avec n'importe quel chien issu de la population de référence' (cette différence peut par exemple être due au fait que dans ce monde, les chiens de prairie entretiennent de fréquents rapports incestueux). D'autre part, Max' a un frère, Félix', avec lequel il partage 50% de gènes par voie de descendance.

Si on assimile *r* au degré de parenté par voie de descendance, on pensera que le coefficient d'apparentement entre Max et Félix et Max' et Félix' est exactement le même. Il est vrai que les deux couples de frères partagent 50% de leurs gènes par voie de descendance. Mais cette manière de concevoir le coefficient d'apparentement ne prend pas en considération la proportion moyenne de gènes communs entre les frères considérés et les individus issus de leur population de référence respective. Or, au niveau du processus de la sélection naturelle, *r* prend une signification différente selon la proportion de gènes communs entre les membres du voisinage social et la population de référence ; plus cette proportion est élevée, moins il vaut la peine pour Max d'agir de manière altruiste envers Félix plutôt qu'envers n'importe quel autre individu de la population de référence ; en effet, en termes absolus, Félix ne partage pas beaucoup plus de gènes avec Max qu'avec n'importe quel individu de la population de référence. En d'autres termes, la force de la sélection de parentèle est dépendante de la proportion de gènes partagés par l'ensemble des membres de la population de référence considérée.

Deuxièmement, si le gène qui induit un comportement altruiste est extrêmement rare, alors le coefficient d'apparentement est pratiquement équivalent au degré de parenté par voie de descendance. En effet, si la probabilité pour qu'un gène soit porté par des individus de la population de référence s'approche de zéro, alors la valeur de *r* dans le calcul de Hamilton, n'est pas influencée (comme dans le modèle des chiens de

prairie) par la proportion de gènes communément partagés dans la population de référence.

Troisièmement, la force de la sélection de parentèle devient très difficile à mesurer si les domaines de compétition et de voisinage social se chevauchent. Notons pour commencer qu'un individu altruiste n'agit pas constamment de manière altruiste. Outre son comportement altruiste et discriminatoire en faveur des ses proches parents, ou plus généralement des individus de son voisinage social, cet individu développe toute une palette d'autres comportements. Parmi ceux-ci, il faut compter des comportements compétitifs, voire agressifs lorsqu'il s'agit d'acquérir des ressources qui lui sont vitales. Il est donc important de faire la différence entre le domaine du voisinage social dans lequel l'individu altruiste distribue ses bienfaits et le domaine de compétition, dans lequel l'individu altruiste ne se comporte précisément pas de manière altruiste. Si le domaine de compétition n'est pas le même que le domaine du voisinage social, la règle de Hamilton peut être appliquée sans difficulté ; l'altruiste distribue ses bienfaits à un certain nombre d'individus et entre en compétition avec d'autres individus. Par contre, si le domaine de compétition chevauche le domaine du voisinage social, dans ce cas on peut s'attendre à ce que les bienfaits du comportement altruiste soient d'autant relativisés que la compétitivité est forte ; l'altruiste distribue ses bienfaits et entre en compétition avec les mêmes individus. En réalité, si le domaine de compétition chevauche le domaine potentiel de voisinage social, il y a fort à parier que la sélection naturelle ne favorise par l'émergence de comportements altruistes (voir TAYLOR 1992 ; QUELLER 1992 ; 1994). Cette difficulté théorique a été testée empiriquement par des chercheurs comme Stuart WEST ou Ashleigh GRIFFIN.[71] Leurs résultats montrent que l'effet de sélection de parentèle sera limité si le domaine de compétition est composé de beaucoup de membres parents.[72] WEST et GRIFFIN font

[71] Ces auteurs ont comparé les comportements et coefficients d'apparentement des guêpes des figues (WEST *et al*. 2001) ainsi que les degrés de prolifération et coefficients d'apparentement de bactéries pathogènes (GRIFFIN *et al.* 2004).

[72] Dans un récent article, Laurent LEHMANN, Nicolas PERRIN et François ROUSSET (2006) montrent que l'effet de la sélection de parentèle garde toutefois sa vigueur si les comportements d'aide n'augmentent pas le degré de compétition à l'intérieur du voisinage social. Cela est possible si les comportements d'aide favorisent l'accès à de nouvelles ressources, par exemple si l'espèce évolue dans un environnement non saturé qui peut encore être colonisé.

également remarquer qu'un des facteurs dont HAMILTON pensait qu'il est susceptible de favoriser la sélection de parentèle, la localisation des individus parents dans un même espace territorial restreint (HAMILTON 1964), est précisément un facteur qui augmente en même temps le taux de compétition locale. Voilà une réalité bien gênante pour déterminer, dans les cas concrets, les effets des forces sélectives opposées dues à l'apparentement et à la compétition… (WEST *et al.* 2001 ; GRIFFIN *et al.* 2004).

Quatrièmement, le domaine d'application de la théorie de HAMILTON est plus large que ce à quoi on pourrait s'attendre.

Tout d'abord, une similitude génétique (en particulier si on prend en considération uniquement un gène à un locus donné) n'est pas forcément corrélée au degré de parenté par voie de descendance. Il est donc possible de comprendre *r* comme une simple mesure de similitude génétique entre deux individus (ou une mesure de probabilité qu'ils partagent un même gène).[73] Dans ce cas, *r* n'est pas forcément dépendant du degré de parenté par voie de descendance ; il peut dépendre d'autres facteurs. Ces autres facteurs sont à caractère pléiotropique, c'est-à-dire qu'un même gène ou un paquet indissociable de gènes agit sur plusieurs caractères à la fois. Par exemple, la tendance à l'action altruiste pourrait être associée à une préférence d'habitat très particulière ce qui permettrait un regroupement automatique des individus altruistes dans un même espace géographique restreint (les individus vivant dans ce micro-habitat formeraient un voisinage social dans lequel on pratiquerait l'altruisme). On pourrait également imaginer que cette tendance à l'action altruiste soit associée à la fois à une marque phénotypiquement observable (des grands yeux, une barbe verte, etc.) et à une tendance à agir de manière discriminatoire en faveur des individus qui présentent cette marque ; il s'agit du fameux phénomène de l'effet « barbe verte ».[74] Dans ce contexte, pour désigner *r*, il ne faudrait alors plus parler de coefficient d'apparentement mais plutôt de coefficient de relation génétique ; dans le premier cas, la valeur de ce *r* élargi correspondrait au rapport entre la probabilité qu'un individu du voisinage social possède le gène responsable du comportement altruiste, relativement à la probabilité

[73] Cette conception étendue de *r* comme simple mesure statistique de similitude génétique (sans prise en compte de l'origine de cette similitude) n'est pas encore présente dans le fameux article de HAMILTON de 1964. Il la propose par contre quelques années plus tard (HAMILTON 1970).

[74] Ce phénomène d'abord formulé par HAMILTON (1964, p. 25) a ensuite été rendu célèbre par DAWKINS (1996/1976, pp. 128-129).

qu'un individu de la population de référence possède ce même gène ; dans le deuxième cas, la valeur de *r* élargi correspondrait à la probabilité avec laquelle un individu altruiste dispense ses bienfaits en faveur d'individus possédant également les gènes qui induisent des comportements altruistes (relativement à la probabilité qu'il dispense ses bienfaits à un tricheur arborant la marque caractéristique de l'altruisme). Ce *r* élargi peut ensuite être utilisé dans la règle de Hamilton puisqu'en fin de compte la sélection naturelle favorise n'importe quel gène capable de générer un grand nombre de copies de lui-même, que ces copies soient transmises par voie de descendance ou non. Ainsi, on ne peut pas associer inextricablement la règle de Hamilton au calcul du degré de parenté par voie de descendance (ce qui est le cas dans le phénomène de la sélection de parentèle).

D'autre part, la règle de Hamilton peut s'appliquer dans des contextes où une population est divisée en différents groupes, et où le comportement altruiste d'un individu affecte de manière égale tous les individus du groupe dont il fait partie (c'est-à-dire de son voisinage social). Ainsi, *r* n'est plus forcément une relation entre deux individus ; il peut également être une relation entre un individu et n'importe quel individu issu du même voisinage social.[75]

ii. La fitness inclusive et le paradoxe de l'altruisme

Comme on le pressent déjà dans les exemples du calcul de la *fitness* inclusive présentés ci-dessus, cette nouvelle définition de la *fitness* nous fait avancer d'un premier pas dans la résolution du problème posé par le paradoxe de l'altruisme. HAMILTON a réfléchi à la question de l'évolution de l'altruisme en adoptant la perspective du gène ; il ne s'est pas demandé comment un individu peut transmettre ce trait à sa progéniture, mais plutôt comment les gènes qui induisent des comportements altruistes peuvent, par ailleurs, se répandre dans l'ensemble du pool génétique d'une population (en dépit du fait que ces comportements sont néfastes pour ceux qui les développent). La réponse à cette question est la suivante : ces gènes peuvent se répandre dans le pool génétique à la condition qu'ils induisent, chez l'individu porteur de ce gène, des comportements altruistes en faveur des individus porteurs de copies d'eux-

[75] Cette extension de la signification du coefficient d'apparentement apparaît chez HAMILTON dans son article de 1975.

mêmes. Ainsi la baisse de viabilité et fécondité de l'individu porteur de gènes qui induisent des comportements altruistes sera compensée par l'augmentation de la viabilité et fécondité des individus bénéficiaires possédant ces mêmes gènes.

Remarquons ici que la sélection des gènes responsables de l'altruisme est fortement dépendante du type de bénéficiaires des actions altruistes : il faut qu'un certain nombre de bénéficiaires partagent ces mêmes gènes. Cela implique deux scénarios possibles :

Dans le premier scénario, l'individu altruiste aide systématiquement ses proches parents au détriment des individus non parents. Cela est possible i) s'il possède à la fois la capacité de reconnaître ses proches parents et celle d'agir de manière discriminatoire en leur faveur ou ii) s'il dispense son altruisme de manière non discriminatoire dans l'environnement qu'il habite et que cet environnement est essentiellement composé d'individus parents ; dans ce cas, il n'est même pas nécessaire que l'agent dispose de la capacité de reconnaître ses proches parents. La parenté est un facteur important car elle garantit la possession d'une grande proportion de gènes en commun via la transmission du matériel génétique par voie de descendance.

Dans le second scénario (nettement moins répandu dans le monde biologique), l'individu altruiste aide avec une plus grande probabilité des individus qui possèdent également les gènes responsables de l'altruisme, même s'ils ne sont pas parents ; cela est possible i) s'il est capable de reconnaître les individus altruistes (via des traits observables qui leur sont propres) et d'agir de manière discriminatoire en leur faveur (il s'agit de l'effet « barbe verte ») ou ii) s'il dispense son altruisme de manière non discriminatoire dans l'environnement qu'il habite et que cet environnement est essentiellement composé d'individus altruistes; dans ce cas, il n'est même pas nécessaire que l'agent dispose de la capacité de reconnaître les individus altruistes.

iii. De la théorie à la vie réelle : les abeilles kamikazes et les marmottes siffleuses

Concrètement, est-il possible que les scénarios mentionnés ci-dessus soient effectivement apparus au cours de l'évolution ?

Le second scénario est hautement improbable, en particulier dans sa première version ; on n'a du moins pas encore observé de tendances à un comportement altruiste qui s'accompagne d'un effet phénotypique facilement observable (si ce n'est les effets

observables des actions altruistes). Je reviendrai plus loin (p. 106) sur les raisons théoriques de ce manque de réalisabilité (voir aussi KELLER & ROSS 1998).[76]

En revanche, le premier scénario est plus réaliste. Par exemple, les mouettes argentées considèrent comme leurs oeufs tous les objets contenus dans le nid qu'elles ont construit ; mettez-y un œuf d'une autre espèce d'oiseau ou même un leurre grossier et elles n'y verront que du feu ! (DAWKINS 1996/1976, p. 145) Certains oisillons considèrent comme leur mère l'individu adulte qu'ils côtoient dans les premiers jours après l'éclosion ; c'est de cette manière que Konrad LORENZ est devenu la « maman » d'une couvée d'oies.[77] Si la plupart des animaux n'ont pas affiné leurs critères de distinction, c'est parce que cela n'était pas nécessaire du point de vue de leur adaptation au milieu dans lequel ils évoluent habituellement ; ce qui importe est que cela fonctionne dans la plupart des cas.

Voyons maintenant si la théorie de HAMILTON permet de rendre compte des cas d'altruisme inexplicables par le biais d'une théorie de la sélection naturelle en termes de *fitness* classique. A cet effet, considérons deux exemples déjà évoqués : celui de l'abeille kamikaze et celui de la marmotte siffleuse.

Pour ce qui est des abeilles, il faut savoir que les individus composant une colonie d'insectes sociaux (fourmis, abeilles, etc.) sont généralement issus de la même mère ; et selon les espèces, un ou plusieurs mâles fournissent le sperme nécessaire à l'activité reproductrice d'une reine. Ainsi, ces colonies se caractérisent par un fort degré de parenté entre les individus qui les composent. Dans un tel contexte, il n'est pas étonnant de constater des comportements hautement altruistes ; si une ouvrière se sacrifie pour

[76] Mêmes les auteurs de cette hypothèse, HAMILTON et DAWKINS, doutent de la possibilité de l'évolution de ce genre d'effets pléiotropiques. Pour un exposé des vertus théoriques et des faiblesses explicatives de l'effet barbe verte, voir DAWKINS 1999/1982, pp. 143-155.

Dans une étude sur la fourmi rouge Solenopsis invicta, Laurent KELLER et Kenneth ROSS sont toutefois parvenus à montrer un cas d'effet barbe verte. Leur découverte est toutefois assez peu révélatrice pour notre propos puisque le gène étudié n'est pas responsable d'un comportement altruiste ; il détermine à la fois pour la production d'une odeur particulière et un comportement meurtrier à l'encontre des femelles à maturité sexuelle qui ne produisent pas cette odeur. A la fin de leur article, Keller et Ross admettent que les effets barbe verte sont extrêmement difficiles à trouver dans le monde animal et proposent une hypothèse (basée sur des considérations génétiques) pour en expliquer la rareté (KELLER & ROSS 1998).

[77] Il s'agit du phénomène de l'« empreinte » expérimenté par LORENZ (1989/1988).

sauver sa communauté, elle contribue à la survie et production d'une multitude de ses frères et sœurs.[78]

Pour expliquer le comportement altruiste des marmottes sentinelles, des biologistes comme Paul SHERMAN et Mark HAUBER se sont intéressés au coefficient d'apparentement entre ces animaux, ainsi qu'au rapport entre le risque pris par les sentinelles et l'avantage qu'en retirent les autres individus du groupe. Selon eux, les résultats obtenus confirment l'hypothèse de la sélection de parentèle de Hamilton ; il s'avère que les sentinelles sont souvent entourés de proches parents (SHERMAN 1977 ; HAUBER & SHERMAN 1998).[79]

Dans cet exemple, on peut toutefois remarquer que le simple calcul de la *fitness* classique pourrait éventuellement suffire à expliquer l'évolution du comportement des marmottes sentinelles. En effet, certaines observations semblent montrer qu'au moyen de ce comportement, les marmottes siffleuses sauvent en priorité leurs propres petits du danger ; de ce fait, on pourrait considérer tous les autres (sœurs, neveux, individus non parents) comme des bénéficiaires non pertinents pour le calcul de la *fitness* (BLUMSTEIN *et al.* 1997). Mais je ne pense pas que cette approche soit judicieuse (même si elle fonctionnait dans le cas présent), car elle fait perdre de vue que les soins parentaux et les comportements d'aide envers d'autres individus parents (par exemple une sœur) sont deux illustrations d'un même mécanisme : la sélection de parentèle. Ce n'est qu'en se détachant de l'individu et en adoptant la perspective du couple gène-phénotype que l'on acquiert une réelle compréhension de l'évolution de ce type de comportement.

Cela dit, on ne peut nier que l'étude du comportement des mammifères s'avère bien plus complexe que celle des insectes ; les mammifères se caractérisent par la

[78] L'explication de la stérilité des ouvrières relève également de la théorie de HAMILTON ; des facteurs environnementaux (hostilité du milieu) combinés aux bénéfices de la socialité et au degré de parenté expliquent pourquoi des jeunes femelles renoncent à leur propre reproduction au profit de celle de leur mère (pour une introduction didactique, voir CHAPUISAT & KELLER 2007; pour une analyse pointue voir BOURKE & FRANKS 1995).

[79] Même si cette explication semble la plus convaincante, il faut savoir qu'il en existe d'autres qui se passent de la sélection de parentèle et invoquent l'avantage individuel retiré par les marmottes siffleuses (TRIVERS 1971). D'autre part, si la théorie de la sélection de parentèle peut être évoquée dans le cas des marmottes, elle n'est d'aucun secours lorsqu'il faut expliquer l'évolution du comportement de sentinelle chez les primates (CHENEY & SEYFARTH 1990).

grande plasticité de leurs comportements si bien que les influences respectives exercées par les gènes, l'environnement et l'expérience sont souvent extrêmement difficiles à déterminer.[80] Dans une certaine mesure, c'est même le cas chez les insectes qui adaptent leur comportement en fonction de certaines données environnementales (voir LEHMANN & PERRIN 2002). Ainsi, pour un comportement donné, si l'on veut déterminer le rôle causal de la sélection de parentèle, il ne suffit pas d'appliquer le calcul de Hamilton ; il faut également prendre en compte d'éventuels paramètres supplémentaires, dont les influences épigénétiques.

iv. La théorie de l'altruisme discriminant

Une nouvelle génération de théoriciens de l'évolution travaille actuellement sur des modèles théoriques capables de rendre compte de la complexité du monde biologique en faisant interagir à la fois les facteurs génétiques et écologiques. J'ai déjà mentionné quelques noms en expliquant que la puissance de la sélection de parentèle est dépendante du chevauchement ou non des domaines de compétition et de voisinage social (p. 60 et suiv.).

Dans la même lignée, Laurent LEHMANN et Nicolas PERRIN (2002) ont élaboré une théorie que l'on pourrait qualifier de « théorie de l'altruisme discriminant ». Sur fond de modèles mathématiques et d'exemples concrets tirés du monde animal, ils montrent que l'altruisme est évolutionnairement plus stable et se propage mieux lorsqu'il est associé à une capacité de discrimination, plus précisément quand les individus peuvent diriger leurs bienfaits en direction d'autres individus apparentés, reconnus comme tels. L'idée n'est pas neuve et a déjà été discutée plus haut, mais l'originalité de leurs travaux vient de ce qu'ils précisent les différentes manières dont la discrimination peut être effectuée et ses effets sur la propagation de l'altruisme. Il est possible de discriminer sur la base d'une ressemblance phénotypique d'origine génétique ou en fonction d'une ressemblance phénotypique d'origine environnementale (par exemple les individus d'une même colonie de fourmis consomment la même nourriture si bien qu'ils dégagent une même

[80] Pour une illustration concrète de ces difficultés, voir l'étude de Jill MATEO sur le comportement des écureuils du sol (2003).

odeur et peuvent se reconnaître entre eux de cette manière)[81] ; ces deux méthodes peuvent être utilisées conjointement. LEHMANN et PERRIN ont montré que certaines conditions favorisent l'évolution d'un type de discrimination plutôt que l'autre ; par exemple, une population composée de petits groupes dont les individus se dispersent peu (ce qui induit un haut degré de parenté génétique entre les membres du groupe) développera des mécanismes de discrimination basés sur la reconnaissance de traits génétiquement déterminés (et inversement). De plus, leurs résultats montrent que la discrimination sur la base de différences dues à l'environnement engendre de véritables comportements sociaux et une coopération à plus large échelle que dans le cas de la discrimination basée sur les traits génétiques.

Ce modèle rend compte du fait qu'au-delà de la simple sélection de parentèle, des facteurs épigénétiques (la manière dont l'environnement affecte les traits des individus, le taux de dispersion des individus) influencent largement l'évolution des comportements d'aide.

Bilan

Au terme de ces réflexions, on comprend comment il est possible que des gènes qui induisent des comportements altruistes puissent se répandre au fil des générations dans le pool génétique d'une population ; cela est possible s'ils induisent généralement des comportements de sacrifice au profit d'individus qui possèdent des copies de ces mêmes gènes ; en d'autres termes, cela est possible si les comportements qu'ils induisent favorisent la propagation de répliques d'eux-mêmes.

Si la sélection de parentèle permet d'expliquer un certain nombre d'observations empiriques, il faut toutefois se garder d'exagérer son importance. On sait que la sélection de parentèle sera faible si la proportion de gènes communément partagés entre le voisinage social et la population de référence est haute. On sait également que la force de la sélection de parentèle peut être réduite à néant si les domaines du voisinage social et de compétition se chevauchent. Et finalement, il ne faut pas minimiser les

[81] Les êtres humains par exemple, discriminent souvent en fonction de traits culturels (les dialectes locaux, les habitudes de comportement, les ornements traditionnels, etc.) qui sont de nature purement environnementale.

influences épigénétiques (principalement lorsque l'on étudie des individus aux capacités cognitives développées).

Dans les faits, l'élégance de la théorie de HAMILTON convaincra les biologistes, choquera des penseurs des sciences sociales (par exemple, SAHLINS 1980/1976) et sera boudée par les philosophes ; il est en effet difficile d'accepter qu'un terme à connotation aussi noble que l'altruisme puisse être expliqué en termes d'avantages et d'égoïsme des gènes. Les penseurs outrés peinent à saisir l'aspect métaphorique de certaines formules utilisées par les sociobiologistes.

Un aspect extrêmement attrayant de la règle de Hamilton est qu'elle permet de contredire le fameux slogan de la survie du plus apte. Ce qui favorise la transmission d'un caractère au fil de l'évolution n'est pas tant le fait qu'il soit bénéfique à l'individu qui le développe. L'important est que les gènes responsables de ce caractère puissent se répliquer efficacement et augmenter en proportion dans l'ensemble du pool génétique.[82]

Pour ce qui est de la controverse autour de l'altruisme évolutionnaire, HAMILTON se situe clairement dans le camp des stratèges de l'égoïsme, ceux qui affirment que tous les comportements altruistes évolutionnaires peuvent être compris en termes d'avantages sélectifs. La question qui se pose maintenant est de savoir si la théorie de HAMILTON est capable de rendre compte de tous les cas d'altruisme que l'on rencontre dans le monde animal. Ce n'est pas l'avis de tout le monde, et c'est ce que nous verrons à la section suivante.

2.2.2. La réciprocité directe

La théorie de la sélection de parentèle permet d'expliquer les comportements altruistes entre individus parents. Or, les biologistes ont pu observer des comportements à première vue altruistes entre individus non parents, voire entre individus d'espèces différentes. En voici deux exemples (qui feront l'objet d'une analyse plus détaillée aux sections 2.2.2.ii et 2.2.2.vi).

[82] « The ultimate criterion which determines whether G will spread is not whether the behavior is to the benefit of the behaver but whether it is to the benefit of the gene G ; and this will be the case if the average net result of the behavior is to add to the gene-pool a handful of genes containing G in higher concentration than does the gene-pool itself. » (HAMILTON 1963, pp. 354-355)

Il existe une espèce de chauves-souris vampires dont les individus qui rentrent repus après une nuit de chasse régurgitent souvent une partie de leur repas dans la gueule des malchanceux rentrés bredouilles ; or on a constaté que ces dons de sang ne se faisaient pas systématiquement en faveur d'individus parents. Comment expliquer ces actions charitables ?

Dans la nature, il existe des symbioses entre individus d'espèces différentes. On en trouve notamment qui lient des grands poissons prédateurs et des petits poissons (ou des crevettes) nettoyeurs qui entrent dans les branchies de leurs hôtes pour se nourrir des parasites qui y résident. Comment expliquer que ces grands poissons ne fassent pas qu'une bouchée des petits et se prêtent même à des postures très inconfortables (qui les rendent vulnérables à des attaques ennemies) pendant les séances de nettoyage ?

Puisque la théorie de HAMILTON ne permet pas de rendre compte de ces cas, il faut trouver une nouvelle explication. C'est dans cette entreprise que se sont lancés les biologistes de l'évolution et quelques théoriciens des jeux.

i. Les conditions de l'évolution de l'altruisme réciproque

Robert TRIVERS est le premier à avoir émis l'hypothèse que les comportements altruistes en faveur de non-parents peuvent évoluer à condition qu'ils s'accompagnent de retours de services ultérieurs de la part des bénéficiaires (TRIVERS 1971). C'est la naissance de la théorie de l'altruisme réciproque ou de la réciprocité ; la réciprocité se caractérise par un délai entre les actions altruistes et un retour de service. Voici l'explication théorique d'un cas de relation de réciprocité réussie : en t_1, l'un des partenaires de l'interaction prend un risque en produisant une action qui a pour effet d'augmenter la *fitness* de l'autre au détriment de sa propre *fitness*. Lorsque l'occasion se présente (en t_2), le receveur retourne le service (ou son équivalent). Dans certains cas, il se peut que t_1 et t_2 se confondent ; pour qu'on puisse réellement parler d'altruisme réciproque dans ces situations, il faut que les deux protagonistes ignorent le comportement choisi par l'autre parti. Une relation de réciprocité tourne court lorsque l'un des deux partenaires refuse de rendre la pareille alors qu'il est en mesure de le faire ou lorsqu'il en est incapable à long terme.

Voici une liste de conditions nécessaires au développement d'une relation de réciprocité entre deux individus :

- Il faut que les individus possèdent une tendance à agir de manière altruiste précisément en faveur d'individus qui pourront leur rendre ultérieurement un service proportionné. Cela est possible s'ils possèdent la capacité de reconnaître et de se souvenir des individus envers lesquels ils ont agi de manière altruiste (mais cela requiert des facultés mentales relativement développées dont beaucoup d'animaux sont dénués) ou s'ils agissent de manière altruiste dans des circonstances extrêmement réglées (par exemple toujours au même lieu et dans le cadre d'un rituel particulier).

- Il faut que les situations de réciprocité apparaissent régulièrement. Cela implique trois conditions : premièrement, que les individus aient l'occasion de se rencontrer régulièrement (c'est le cas par exemple s'ils vivent ensemble dans un espace géographique restreint ou s'ils sont membres d'un petit groupe itinérant) ; deuxièmement que les individus aient une espérance de vie suffisante pour pouvoir interagir un bon nombre de fois ; troisièmement que l'intervalle de temps entre le moment où un individu effectue une action altruiste et le moment où il reçoit le service en retour ne soit pas trop grand.

- Il faut que les rôles s'inversent régulièrement, c'est-à-dire que le receveur devienne le donneur et inversement.

- Il faut que le gain, pour le receveur, soit supérieur au coût, pour le donneur. Par exemple, si un individu possède une grande réserve de nourriture, le fait d'en donner une petite partie à un individu affamé lui coûte peu tout en étant un gain très appréciable pour le receveur.

- Il faut que les partenaires de l'interaction soient exposés de manière plus ou moins symétrique aux interactions altruistes ; en d'autres termes, les coûts versus gains doivent être proportionnés pour les différents protagonistes. Par exemple, si un singe prend du temps à épouiller un autre singe afin que ce dernier partage sa nourriture avec lui, l'énergie investie dans l'épouillage en comparaison du gain en nourriture doit correspondre plus ou moins à l'investissement du don de nourriture en comparaison avec le gain du toilettage. La réciprocité échoue si les coûts et gains ne peuvent pas être proportionnés. Par exemple, dans une troupe de singes très hiérarchisée, le chef qui bénéficie de tous les privilèges et un jeune singe freluquet du bas de la hiérarchie ne sont pas exposés de manière symétrique aux interactions altruistes (puisque le premier

peut de toute façon se servir de tous les biens qu'il convoite) ; une relation de réciprocité ne peut donc pas être instaurée entre ces deux individus.

ii. De la théorie à la vie réelle : l'exemple d'une symbiose de nettoyage

Voyons si ce cadre théorique permet d'éclairer la symbiose de nettoyage dont il est question plus haut. On trouve beaucoup de symbioses de nettoyage chez les poissons. Dans son article de 1971, Robert TRIVERS en présente un exemple. Les protagonistes sont d'une part des gros poissons prédateurs qui se nourrissent de petits poissons, d'autre part, des petits poissons ou des crevettes qui se nourrissent de parasites collés à la surface des gros poissons ; les premiers sont les hôtes, les second les nettoyeurs. Les observations montrent que les hôtes prennent garde de ne pas manger les nettoyeurs. Ces derniers, pour se faire reconnaître (et ne pas finir dans l'estomac de leurs hôtes), demeurent toujours dans le même site et arborent des couleurs et des comportements reconnaissables. D'autre part, les hôtes changent de couleur ou de comportement lorsqu'ils ont besoin d'être nettoyés et signalent aux nettoyeurs lorsqu'ils ne désirent plus être toilettés (afin que ces derniers aient le temps de se mettre en sécurité). Enfin, durant le toilettage, les hôtes adoptent des positions inconfortables, les rendant vulnérables aux éventuelles agressions ennemies.

TRIVERS considère cette symbiose comme un cas d'altruisme réciproque car les hôtes s'abstiennent d'avaler les nettoyeurs. Il est vrai que la plupart des conditions nécessaires au développement d'une relation de réciprocité sont remplies : ritualisation de l'interaction (à défaut de reconnaissance individuelle), occurrence régulière des situations d'interaction, symétrie de l'interaction (l'hôte y gagne en étant débarrassé de ses parasites et le nettoyeur y gagne en se remplissant la panse) et rapport gain / coût (le nettoyeur prend le risque de se faire manger par erreur et l'hôte se retient de manger tout en adoptant des postures inconfortables et parfois dangereuses). La symbiose de nettoyage peut toutefois être considérée comme un cas limite d'altruisme réciproque en raison du fait que le temps de l'investissement n'est pas différé par rapport à celui du retour de service ; le nettoyage et le nourrissage se font de manière simultanée si bien qu'on ne peut pas réellement parler d'inversion des rôles du donneur et du receveur. Ainsi on peut se demander si les symbioses entre individus d'espèces différentes peuvent réellement être considérées comme des cas d'altruisme réciproque.

Quoi qu'il en soit, l'analyse de TRIVERS trouvera une confirmation empirique si l'on peut montrer qu'il vaut réellement la peine pour l'hôte, de préserver la vie du nettoyeur plutôt que de le manger (si ce n'était pas le cas, on pourrait vraiment parler d'altruisme pur de la part de l'hôte !). Pour cela, il faut montrer empiriquement que les hôtes souffrent effectivement de parasitisme, qu'il leur est difficile et assez périlleux de se lancer chaque fois à la recherche d'un nouveau nettoyeur, que les nettoyeurs qui ont déjà exercé sur un hôte sont faciles à retrouver, qu'ils vivent suffisamment longtemps pour répéter le nettoyage et que les hôtes réutilisent effectivement les mêmes nettoyeurs dans la mesure où les conditions sont favorables. Dans son article de 1971, TRIVERS présente des résultats empiriques qui confirment son analyse.[83]

iii. La théorie des jeux et l'altruisme réciproque

La théorie de la réciprocité est fortement liée au calcul des coûts et des intérêts des individus qui interagissent ainsi qu'à l'alchimie de stratégies comportementales concurrentes dans un milieu donné. Très vite, des théoriciens des jeux[84] se sont intéressés à ces questions et ont développé des modèles destinés à simuler des environnements sociaux compétitifs, des stratégies comportementales utilisables dans ces environnements et l'effet de la sélection naturelle sur ces stratégies comportementales (MAYNARD SMITH & G. PRICE 1973 ; AXELROD 1996/1984). Il s'agissait non seulement de tester la robustesse et la stabilité de différentes stratégies

[83] Dans un article plus récent, Redouan BSHARY et Daniel SCHÄFFER (2002) indiquent toutefois que la plupart des symbioses de nettoyage ne fonctionnent pas de manière symétrique (comme c'est le cas dans l'exemple de TRIVERS) car les hôtes ne sont souvent pas des prédateurs et se nourrissent de végétaux. Pour expliquer ces situations asymétriques, les auteurs font appel à la théorie du marché biologique (que je présenterai plus loin ; p. 85). Les cas de symbioses entre nettoyeurs et poissons non prédateurs sont toutefois moins intéressants pour notre propos puisque l'aspect apparemment altruiste des hôtes n'apparaît pas.

[84] La théorie des jeux a été développée par des mathématiciens au début du siècle afin de comprendre les systèmes économiques et militaires (BOREL 1921 ; VON NEUMANN 1928 ; VON NEUMANN & MORGENSTERN 1944). L'objectif plus général de cette théorie est de savoir quelles stratégies sont avantageuses dans un contexte où les décisions des uns ont des conséquences sur les autres. Une branche particulière, la théorie des jeux évolutionnaires (qui nous intéresse ici) a été créée par le statisticien Ronald FISHER (1939) et le biologiste John MAYNARD SMITH (1982). Elle s'emploie à analyser les stratégies évolutionnaires des organismes et des espèces.

dans des milieux plus ou moins hostiles, mais également de préciser les conditions nécessaires à l'émergence de l'altruisme réciproque.

Illustrons la manière dont on procède dans le cadre de ce type de recherches au moyen d'un exemple connu ; celui de l'épouillage mutuel.

1. On commence par la présentation du milieu dans lequel évoluent les acteurs de l'interaction. Pour notre exemple, admettons que les individus d'une espèce d'oiseaux soient parasités par des tiques. Il est important pour ces oiseaux de se débarrasser régulièrement de leurs parasites sans quoi ils risquent de contracter une maladie mortelle. Chaque individu peut s'épouiller partout sauf au sommet de sa tête, lieu que les tiques affectionnent tout particulièrement ; la seule manière de se débarrasser de ces tiques récalcitrantes est de se laisser épouiller par un autre oiseau. Supposons également que toutes les conditions nécessaires à la réciprocité évoquées précédemment (p. 72) soient réalisées (capacité de reconnaissance, occurrences régulières des situations d'interaction, inversion des rôles, rapport gain/coût et symétrie de l'interaction). Voilà d'excellentes conditions pour le développement d'un comportement altruiste réciproque !

2. Dans une deuxième étape, on pose les stratégies comportementales possibles ; celles-ci sont génétiquement programmées et les individus les appliquent à la lettre durant toute leur vie. Dans notre exemple, imaginons qu'il existe deux stratégies possibles chez les oiseaux parasités : la stratégie E (Epouilleur) qui consiste à épouiller son voisin chaque fois que l'occasion se présente et la stratégie T (Tricheur) qui consiste à ne jamais épouiller son voisin.

3. Dans un troisième temps, il s'agit de simuler une longue série d'interactions et après chaque interaction, calculer les gains et les coûts des acteurs. Il est également possible de simuler l'évolution de différentes générations en établissant un taux de reproduction calculé en fonction de l'ensemble des gains récoltés par chaque individu à la fin de sa vie ; si, au terme de sa vie (une vie correspond à un nombre *x* d'interactions), le résultat cumulé des interactions d'un individu se solde par un bilan coûts/gains positif, alors on lui attribuera une grande descendance (et inversement si le bilan est négatif). Revenons à notre exemple et imaginons qu'un individu A Epouilleur, épouille la tête d'un individu B Tricheur. Lorsque vient le tour de B, celui-ci refuse d'épouiller A et s'en va. Résultat : l'individu A a investi de l'énergie et n'a reçu aucun service en retour ; l'individu B tire bénéfice de l'épouillage et ne paie aucun coût en

échange. En simulant l'évolution probable de ce type d'interactions, on obtient une situation typique du paradoxe de l'altruisme : au fil des générations, la stratégie comportementale E va disparaître sous la pression de la sélection naturelle.[85] On constate que la combinaison de stratégies E et T ne mène pas à une situation d'altruisme réciproque ; en réalité on se retrouve devant le cas de figure de l'évolution de l'altruisme (stratégie E) versus égoïsme (stratégie T) vu précédemment.

4. Dans une quatrième étape, on répète plusieurs fois l'opération en variant les stratégies utilisées dans le milieu ainsi que les proportions dans lesquelles ces stratégies sont représentées au début du jeu. Dans notre exemple, pour que la réciprocité soit rendue possible il faut ajouter une nouvelle stratégie, R (Rancunier), qui consiste à épouiller tout individu rencontré pour la première fois, et à refuser d'épouiller tout individu qui a refusé d'épouiller lors de la dernière rencontre. La simulation des interactions de E, T et R nous prédit que la stratégie R, si elle est bien représentée au départ dans l'ensemble de la population, permet de maintenir la réciprocité ; les Rancuniers se feront des amis Epouilleurs et refuseront systématiquement de perdre du temps et de l'énergie en toilettant les Tricheurs qui leur ont déjà refusé ce service ; cette fois, ce sera à la population des Tricheurs de décliner (sans forcément disparaître complètement). Si par contre, au début du jeu, les Rancuniers sont trop rares par rapport aux Tricheurs, ils feront trop souvent de mauvaises « premières rencontres » et subiront le même sort que les Epouilleurs ; et avec eux disparaîtra la réciprocité.

5. La dernière étape consiste à tirer les conclusions. Grâce à l'exemple théorique de l'épouillage mutuel, on sait qu'une bonne stratégie altruiste réciproque se caractérise par la sympathie lors de la première rencontre et la rancune en cas de tricherie (les Tricheurs sont punis lors de la prochaine rencontre). C'est Robert AXELROD qui a découvert les vertus de cette stratégie altruiste réciproque. Je reviendrai sur ses recherches, mais j'aimerais au préalable présenter une notion très commode pour l'étude de l'évolution de stratégies comportementales : celle de stratégie évolutionnairement stable.

[85] En fait, peu importe la distribution de la population, les Tricheurs s'en sortiront toujours mieux que les Epouilleurs, et cela même si toute la population est vouée à l'extinction (ce qui est le cas dans notre exemple).

iv. La stratégie évolutionnairement stable

La notion de stratégie évolutionnairement stable (SES) a été élaborée par John MAYNARD SMITH (1973 ; 1982). C'est un concept théorique extrêmement intéressant car il peut servir de mesure d'efficacité de différentes stratégies comportementales.

Une SES est une stratégie qui, si elle est adoptée par la plupart des individus dans un milieu donné, ne peut être supplantée par aucune autre stratégie présente dans ce milieu. Une SES peut également être une combinaison de différentes stratégies (par exemple, l'individu applique la stratégie A dans 2/3 des interactions et la stratégie B dans 1/3 des interactions) ; dans ce cas on parle de stratégie mixte. En guise d'illustration, reprenons l'exemple de l'épouillage mutuel et plaçons-nous au niveau de la dynamique des populations.

Si dans le milieu de départ, on trouve une répartition inégale des stratégies T (Tricheur), E (Epouilleur) et R (Rancunier) de sorte qu'il existe beaucoup de T et très peu de R et de E, au fil des générations, la sélection se chargera de la disparition de l'ensemble des R et des E. Dans se cas, on peut dire que T est une SES (et cela même si elle conduit à l'extinction de la population par suite de maladies causées par les tiques !).

Si dans le milieu de départ, on trouve une répartition homogène des stratégies T, E et R de sorte que E et R soient aussi bien représentées que T, la sélection à l'œuvre au fil des générations défavorisera T (très rapidement, les Tricheurs se feront punir par les Rancuniers) et E (les Epouilleurs seront exploités par les Tricheurs) au profit de R. Il est très probable que E disparaisse complètement. En revanche T se portera assez bien dès lors qu'elle sera représentée par un petit nombre d'individus ; en effet, moins les Tricheurs sont nombreux, plus ils ont de chance de rencontrer un Rancunier donné pour la première fois ; dans ces cas, ils sont servis ! Ainsi, on se retrouve avec une configuration mixte évolutionnairement stable, composée d'une grande majorité de Rancuniers et d'une petite proportion de Tricheurs.[86]

[86] Si cette explication est suffisante pour les besoins de cet ouvrage, elle paraîtra excessivement simplificatrice aux yeux d'un spécialiste de la théorie des jeux. Pour être précis, il faudrait établir une distinction plus claire entre le processus de la dynamique des réplicateurs et les SES. Techniquement parlant, une SES n'est qu'*une seule* stratégie, même si elle peut être mixte, c'est-à-dire que chaque individu de la population joue une stratégie qui est une combinaison linéaire de plusieurs stratégies à disposition. La stabilisation d'une coexistence de plusieurs stratégies au sein de la population

Quelle que soit la répartition de départ des stratégies, E ne sera jamais une SES ; il suffit d'introduire un seul Tricheur dans une population composée entièrement d'Epouilleurs pour déstabiliser E à plus ou moins long terme.

Par ces exemples, on voit qu'une SES présente une certaine perfection : elle ne peut pas être supplantée par une stratégie existant dans l'environnement dans lequel elle s'est stabilisée.[87] Cette perfection trouve néanmoins ses limites. Premièrement, la stabilité d'une stratégie n'implique pas forcément qu'elle soit bonne pour les individus qui la pratiquent (dans l'exemple de l'épouillage mutuel nous avons vu que si T devient une SES, les individus seront affaiblis par la maladie pour cause de manque de soins). Deuxièmement, dans monde biologique, une SES peut être déstabilisée par l'apparition d'une nouvelle stratégie mutante (qui n'était pas présente auparavant dans le milieu dans lequel la SES s'est imposée). Troisièmement une SES peut disparaître si les conditions de l'environnement changent.

Notons qu'il existe un lien étroit entre la SES et l'adaptation biologique : dans la nature, une SES qui ne mène pas à sa propre destruction peut être considérée comme une stratégie adaptée à l'environnement dans lequel elle s'est stabilisée.

v. Le dilemme du prisonnier itératif

Afin de définir les conditions d'émergence et de stabilisation (au sens de SES) des comportements coopératifs, Robert AXELROD (1996/1984) a modélisé sur ordinateur le jeu du dilemme du prisonnier itératif et un grand nombre de stratégies utilisables dans ce jeu. Comme nous le verrons, l'intérêt du dilemme du prisonnier itératif est qu'il permet de simuler une situation de sélection naturelle de stratégies comportementales.

AXELROD procède de la façon décrite ci-dessus avec le modèle des épouilleurs.

(polymorphisme stable) relève en revanche de la dynamique des réplicateurs. Ainsi l'ensemble des SES est un sous-ensemble des états d'équilibre de la dynamique des réplicateurs, caractérisé par le fait qu'une seule stratégie (simple ou mixte) s'impose dans la population.

[87] Au niveau de la dynamique des populations, une autre vertu d'un équilibre stable est que toutes les stratégies présentes dans l'environnement ont la même *fitness*.

1. Il commence par la présentation du milieu dans lequel évoluent les acteurs de l'interaction : il s'agit d'une situation de dilemme du prisonnier itératif qui se présente comme suit. Chaque joueur utilise du début à la fin du jeu une seule et même stratégie définie au départ. Le jeu se déroule en un grand nombre de coups. A chaque coup, les joueurs se rencontrent deux par deux. Lors de ces rencontres entre deux joueurs, chacun a le choix entre deux actions : coopérer ou faire défection. Le choix de chaque joueur est dicté par la stratégie définie au départ (par exemple, la stratégie « coopère une fois sur deux ! »). Lors d'une rencontre entre deux joueurs, aucun ne peut connaître, avant d'avoir joué son coup, l'action décidée par son adversaire. Par contre, chaque joueur peut se souvenir du dernier coup joué par un adversaire lors d'une précédente rencontre. Après chaque rencontre, les points sont répartis comme suit :[88]

		Joueur B	
		coopération	*défection*
Joueur A	*coopération*	3 / 3	0 / 5
	défection	5 / 0	1 / 1

Jusqu'à la fin du jeu, les joueurs cumulent les points récoltés à chaque rencontre. Le vainqueur est celui qui a récolté le plus grand nombre de points.

2. Dans une deuxième étape, AXELROD pose les stratégies comportementales possibles. A cet effet, il a prié un grand nombre de scientifiques du monde entier de développer la stratégie qui leur paraît la meilleure. A celles-ci (dont certaines sont extrêmement complexes et font entrer des calculs de probabilités), il a ajouté quelques stratégies simples comme « choisis au hasard ! ».

3-4. Après avoir transcrit toutes les stratégies en code informatique, AXELROD a simulé par ordinateur un tournoi qui consiste en une longue série d'interactions au cours desquelles les différentes stratégies se confrontent entre elles et également à elles-mêmes. Il a ensuite répété plusieurs fois l'opération en variant les stratégies utilisées dans le milieu ainsi que les proportions dans lesquelles ces stratégies sont représentées au début du jeu. La stratégie qui a largement gagné le plus de tournois s'appelle *Donnant Donnant* (Tit For Tat). Etonnamment, elle est à la fois coopérative et d'une simplicité désarmante. Elle se compose de deux règles : a/ « coopère toujours lors d'une

[88] Il est possible d'imaginer d'autres répartitions des points ; l'important est de maintenir les mêmes rapports entres ces différents gains.

première rencontre ! », et b/ « copie l'action précédente de ton adversaire ! » (s'il a coopéré la dernière fois que tu l'as rencontré, coopère ; et inversement).

5. Enfin, il tire les conclusions. Au premier abord, on pourrait penser que *Donnant Donnant* n'est pas une bonne stratégie puisqu'elle est coopérative. En effet, quel que soit le choix de l'adversaire, en coopérant, on ne peut pas obtenir de meilleur résultat que lui : si je coopère et qu'il fait défection, j'obtiens 0 points et lui 5 ; si je coopère et qu'il coopère également, nous obtenons tous les deux 3 points.

La raison pour laquelle *Donnant Donnant* s'avère une bonne stratégie tient au fait qu'elle est utilisée dans un jeu itératif. Le fait qu'un grand nombre de coups soient joués et que les différents gains s'additionnent changent la donne : il ne s'agit pas de gagner à chaque coup contre un joueur qui sera ensuite éliminé du jeu (comme c'est le cas dans des jeux à somme nulle) mais d'accumuler le plus de points possibles au fil des rencontres. Dans ces conditions, les interactions coopératives portent leurs fruits (trois points à chaque rencontre).

C'est pourquoi, une stratégie coopérative est efficace à long terme à condition qu'elle ne soit pas trop souvent exploitée par l'autre. Or c'est précisément le cas de *Donnant Donnant*. Un joueur qui développe une telle stratégie sera toujours coopératif lors d'une première rencontre ainsi qu'avec tous les individus qui ont coopéré avec lui lors de leur précédente rencontre ; d'autre part, il se préserve assez bien de l'exploitation en étant rancunier avec les adversaires qui n'ont pas coopéré lors de leur précédente rencontre.[89] Enfin, *Donnant Donnant* est indulgent puisqu'il rétablit la coopération dès que son adversaire se remet à coopérer.

L'intérêt de la recherche d'AXELROD réside dans le fait que la stratégie *Donnant Donnant* correspond précisément à l'altruisme réciproque défini par TRIVERS.[90] Cette constatation a donné l'idée à AXELROD de tester la stabilité évolutionnaire de la stratégie *Donnant Donnant*. Pour ce faire, il a développé une simulation par ordinateur d'un contexte de sélection naturelle : il a organisé une suite de tournois, la configuration

[89] Notons que la stratégie altruiste pure « coopère toujours! » est non seulement instable (elle se fait régulièrement exploiter par les stratégies non coopératives) mais également néfaste pour l'évolution de la coopération dans une société puisqu'elle permet aux stratégies non coopératives d'accumuler des gains sans jamais subir la punition d'une non-coopération en retour (AXELROD 1996/1984, p. 129 ; MACKIE 1989).

[90] Remarquons également que l'exemple des épouilleurs est une forme de dilemme du prisonnier.

de départ de chaque nouveau tournoi reflétant les résultats du précédent ; c'est-à-dire que les stratégies qui ont obtenu de bons scores au terme d'un tournoi se trouvent ensuite mieux représentées (en termes de nombre de joueurs qui appliquent ces stratégies) dans le tournoi suivant que celles qui ont obtenu de mauvais scores.

Dans les faits, la stratégie *Donnant Donnant* est parvenue à un équilibre évolutionnairement stable dans un grand nombre de configurations testées ; il suffit qu'au départ, une proportion minimale[91] d'individus coopérateurs soit intégrée pour qu'au fil des tournois, *Donnant Donnant* obtienne une très haute probabilité d'évincer les autres stratégies jusqu'à atteindre une bonne stabilité évolutionnaire.[92] Précisions ici que les résultats obtenus n'indiquent rien sur les origines évolutionnaires de la stratégie *Donnant Donnant*, c'est-à-dire sur les facteurs qui permettent l'apparition et la propagation de *Donnant Donnant* jusqu'à ce seuil minimal.

Appliquer la théorie d'AXELROD au monde biologique revient à imaginer que les stratégies sont des types de comportements génétiquement déterminés ; ainsi, une stratégie serait l'effet phénotypique d'un gène. D'autre part, la théorie d'AXELROD montre que pour qu'un comportement altruiste réciproque puisse être sélectionné, il faut qu'à long terme, il soit favorable à l'individu altruiste réciproque (et par là même aux gènes qui codent pour son comportement). C'est le cas si les conditions suivantes sont réalisées : il faut une bonne probabilité que les individus d'une population se rencontrent un bon nombre de fois au cours de leur vie dans des circonstances où le coût

[91] Toutefois, dans un monde de défection inconditionnelle, quelques individus adoptant la stratégie *Donnant Donnant* ne pourront prospérer. En effet, ils ne rencontreront pas suffisamment de partenaires coopératifs. Dans un tel monde, c'est la stratégie « Fais toujours défection ! » qui deviendra évolutionnairement stable.

[92] Il convient de remarquer que même si *Donnant Donnant* arrive à un équilibre stable, il se peut que cet équilibre soit mixte et admette la persistance d'une petite proportion de stratégies non coopératives ; des individus non coopérateurs peuvent survivre dans un monde de *Donnants Donnants* s'ils ont de bonnes chances de rencontrer régulièrement des individus *Donnant Donnant* pour la première fois (et dans ce cas, obtenir le gain de la défection : 5 points). D'autre part, AXELROD est probablement un peu trop optimiste au sujet de la stabilité évolutionnaire de *Donnant Donnant* ; Robert BOYD et Jeffrey LORBERBAUM (1987) ont pu montrer que dans des conditions de mutation particulières, la stabilité de *Donnant Donnant* peut être ébranlée.

potentiel de la coopération ne soit pas trop élevé[93] ; d'autre part, il faut que les altruistes réciproques soient capables de reconnaître les individus avec lesquels ils ont déjà interagi et de se souvenir des comportements précédemment adoptés par ces individus.

Pour expliquer comment, dans un monde biologique, la stratégie *Donnant Donnant* peut apparaître et se développer jusqu'au seuil critique qui lui permettra ensuite de se propager et de devenir évolutionnairement stable, AXELROD fait appel à HAMILTON. Selon ces deux auteurs, « il est possible d'imaginer que les bénéfices de la coopération dans des situations analogues au dilemme du prisonnier peuvent commencer à être récoltés par des groupes de proches parents » (AXELROD & HAMILTON 1996/1984, p. 95). Evidemment, les auteurs font référence ici au phénomène de sélection de parentèle. Ils poursuivent en expliquant qu'on pourrait imaginer une situation dans laquelle les comportements altruistes en faveur des proches parents s'étendent aux individus dont le degré de parenté est de moins en moins certain. Dans un tel environnement à la fois coopératif et incertain, un mutant altruiste réciproque s'en sortira bien[94] et génèrera une ligne de descendants qui, peu à peu, pourront s'imposer au-delà de ce contexte restreint.[95]

vi. De la théorie à la vie réelle : la réciprocité chez les vampires

Les résultats de la théorie des jeux évolutionnaires ont pu être observés dans le monde animal. L'éthologue Gérald WILKINSON et ses collaborateurs ont mené un travail d'observation sur le comportement de chauves-souris vampires *Desmodus rotundus* (WILKINSON 1984 ; 1990). Leurs observations révèlent que des vampires non apparentés se nourrissent entre eux. Chaque nuit, les vampires sortent de leur caverne à la recherche de victimes (généralement chevaux ou vaches) dont ils boivent le sang. Entre 7 et 30% d'entre eux rentrent bredouille de leur chasse. Les vampires ne peuvent pas survivre à un jeûne de plus de 3 jours. Les observations ont montré que les individus qui

[93] Par exemple, si la répartition des points en cas de coopération de l'un et défection de l'autre était de 0/10 au lieu de 0/5 (la répartition des autres configurations restant la même), même *Donnant Donnant* ne pourrait pas s'avérer une stratégie efficace étant donné le trop grand risque lié à la coopération.

[94] A ce propos, précisons que l'altruisme réciproque peut aussi bien être pratiqué entre individus parents qu'entre individus non parents.

[95] Ce raisonnement est très similaire à celui de John TOOBY et Leda COSMIDES (voir sections 2.3.6 et 3.4.2).

rentrent repus de sang à la caverne offrent régulièrement une partie de leur repas aux malheureux affamés en régurgitant du sang. WILKINSON et collègues ont alors cherché à comprendre la logique de ce don de sang et surtout à déceler s'il s'agit d'un cas de sélection de parentèle ou d'altruisme réciproque. Les résultats indiquent que les échanges de sang entre vampires résultent simultanément d'une sélection de parentèle et d'un altruisme réciproque, car 30% des dons de sang s'effectuent entre individus non parents (les 70% restants sont des dons des mères à leur progéniture).

L'étude de WILKINSON et collègues montre que presque toutes les conditions de la réciprocité sont remplies dans le cas du don de sang chez les vampires. Les donneurs semblent capables d'identifier les individus susceptibles de leur rendre la pareille et aider ceux-ci de préférence. En effet, les observations ont montré que les vampires sont capables de se reconnaître par identification sonore et qu'il existe une corrélation entre le toilettage mutuel et le don de sang ainsi qu'une préférence pour les individus demandeurs qui ont précédemment donné du sang (ce qui indique que les rôles de donneur et de receveur alternent régulièrement). D'autre part, les partenaires de l'interaction ont régulièrement affaire les uns aux autres puisqu'ils passent souvent leurs nuits dans la même grotte ou tronc d'arbre creux ; cette situation permet un échange régulier des rôles de donneur et receveur. De plus, ces échanges ne se font pas uniquement entre individus parents, puisque WILKINSON et collègues ont pu déceler des associations durables de femelles non consanguines. Enfin la condition selon laquelle le gain pour le receveur doit être supérieur au coût pour le donneur est réalisée : les observations ont montré que les vampires ne donnent du sang qu'aux individus dont il reste moins de 24 heures de réserve et que lors d'un don de sang, les receveurs gagnent plus de 18 heures de survie alors que le coût pour le donneur engendre une perte d'énergie vitale d'environ 3 heures.[96]

vii. Les nouvelles théories de la réciprocité directe

Les travaux de MAYNARD SMITH et AXELROD portant sur la théorie des jeux évolutionnaires ont exercé un impact considérable sur les études du comportement social et sont encore aujourd'hui largement cités dans la littérature. Beaucoup de chercheurs ont tenté d'affiner le modèle du dilemme du prisonnier itératif développé par

[96] Nous verrons toutefois (p. 86) que l'interprétation de ces données peut être remise en question.

AXELROD. C'est le cas de Gilbert ROBERTS et Thomas SHERRATT (1998) qui ont modifié les conditions de jeu du dilemme du prisonnier de manière à le rendre plus proche de la réalité. Selon eux, le modèle développé par AXELROD n'est pas réaliste puisqu'il ne laisse aux joueurs que deux choix possibles : coopérer ou faire défection. Or, dans la réalité, la coopération est rarement de l'ordre du tout ou rien ; il existe des formes de défection plus subtiles comme un investissement coopératif légèrement inférieur à celui de son partenaire (TRIVERS avait déjà traité cette question dans son article de 1971). Les auteurs montrent que si l'on intègre cette stratégie comportementale dans la modélisation théorique, on constate qu'elle aura pour effet d'éroder lentement la coopération dans l'ensemble de la population (et cela malgré une bonne représentation de départ de la stratégie *Donnant Donnant*). Suite à ces considérations, G. ROBERTS et SHERRATT proposent une nouvelle stratégie capable de contrer les défections subtiles : ils la nomment « augmente les enjeux ! » (*raise the stakes*) ; il s'agit de coopérer un minimum au départ (en proposant par exemple un tout petit service comme deux minutes d'épouillage) et si la pareille est rendue, augmenter ensuite l'investissement de départ. Les auteurs montrent que cette nouvelle stratégie permet de maintenir la coopération dans une population malgré la pratique d'investissements différentiels. D'autre part, une fois mise en place, cette stratégie s'avère évolutionnairement stable.

Les travaux de G. ROBERTS et SHERRATT se dirigent donc vers un traitement plus réaliste de la réciprocité. Mais cela n'est pas encore suffisant ! D'une part, ils ne proposent pas d'exemples concrets d'application de la stratégie « augmente les enjeux ! » dans le monde animal ; en réalité, le phénomène de l'augmentation des investissements postulé par leur stratégie « augmente les enjeux ! » n'a pas pu être observé (BARRETT *et al.* 1999). D'autre part, leur modèle postule encore un certain nombre de conditions qui ne reflètent pas correctement les circonstances de la sélection naturelle. Dans les faits, les animaux peuvent choisir leurs partenaires d'interactions (développer des préférences pour certains types de partenaires plutôt que d'autres) ; or ce facteur n'est pas pris en compte par G. ROBERTS et SHERRATT (pas plus que par AXELROD d'ailleurs).[97] Il serait également intéressant d'introduire dans la modélisation,

[97] HAMMERSTEIN fait remarque que dans les modèles de théorie des jeux, les joueurs sont forcés de jouer avec les partenaires qui leurs sont assignés : « Players are treated as if they were attached to each other by some 'magic glue' » (HAMMERSTEIN 2003b, p. 84). Or cela ne correspond pas à la réalité.

le paramètre de la mobilité des individus car il s'agit d'un facteur qui empêche la stabilité de l'altruisme réciproque (ENQUIST & LEIMAR 1993).

Les travaux récents de Ronald NOË et Peter HAMMERSTEIN (1994) vont dans le sens d'un traitement plus réaliste de la réciprocité en introduisant le paramètre du choix des partenaires de coopération. Leur théorie du marché biologique (*biological market theory*), se veut une alternative aux modèles d'altruisme réciproque. Cette théorie met l'accent sur la formation de partenariats entre animaux qui s'accordent sur un échange de services. Selon NOË et HAMMERSTEIN, dans les circonstances naturelles dans lesquelles l'évolution prend place, les individus n'entrent généralement pas en contact interactif par pur hasard (ce qui est postulé dans les modèles classiques de théorie des jeux itératifs). La plupart du temps, ils ont le choix entre plusieurs partenaires. Ils peuvent également décider à tout moment de ne plus coopérer avec un individu avec lequel ils ont interagi une série de fois et changer de partenaire. Ainsi, de nouveaux paramètres entrent en jeu : celui du choix des partenaires, celui du moment idéal pour changer de partenaire (cela dépend de la capacité du partenaire à rendre les services rendus, de la quantité de partenaires potentiels disponibles, du coût engendré par un changement de partenaire, du risque de tomber sur un nouveau partenaire peu fiable, etc.). En d'autres termes, la réciprocité fonctionne selon un modèle de marché d'offre et de demande où la possibilité de choisir les partenaires d'interaction et les facteurs liés à ce phénomène doivent être pris en compte.

Outre la théorie du marché biologique et les développements en théorie des jeux, d'autres modèles ont été développés pour rendre compte des comportements partiellement sacrificiels (dans des circonstances particulières) en faveur d'individus non parents. Il y a par exemple la théorie de la pseudo-réciprocité où l'on montre qu'il peut valoir la peine d'aider un autre individu si le bénéfice occasionné a pour effet dérivé d'être favorable à l'individu aidant ; l'idée est que si un organisme bénéficie d'une manière ou d'une autre de la présence d'un autre organisme, un investissement augmentant les chances de survie de ce dernier pourrait être adaptatif (LEIMAR & CONNOR 2003). Le champ d'investigation de la théorie de la réciprocité est loin d'être clos.

viii. L'altruisme réciproque : une rareté

Après 30 ans de recherches sur l'altruisme réciproque, force est d'admettre que peu d'exemples issus du monde animal ont été identifiés (HAMMERSTEIN 2003b).

Il y a bien des exemples de symbioses de nettoyage mais nous avons vu qu'il n'était pas évident de les classer dans la catégorie de l'altruisme réciproque puisqu'il n'y a pas vraiment d'alternance de services.

Quant aux travaux de WILKINSON sur le don de sang chez les vampires (1984 ; 1990), ils ont récemment fait l'objet de critiques. WILKINSON n'a pas de preuve absolue que les vampires appliquent réellement la stratégie comportementale *Donnant Donnant*, car il n'a pas pu prouver empiriquement l'aspect punitif consécutif à une non-coopération ; il n'a pas pu montrer que si un vampire ne donne pas de sang, il sera ensuite puni par un refus lorsqu'il se trouvera lui-même dans une situation de nécessité. Or pour qu'un vampire puisse appliquer une stratégie punitive à la *Donnant Donnant*, il doit posséder la capacité de reconnaître les anciens partenaires d'interaction, se souvenir de leurs actions passées et les classer dans les catégories d'actions coopératives ou non coopératives. Cela requiert des aptitudes mentales extrêmement développées. Or de plus en plus de données empiriques indiquent que ces capacités sont extrêmement rare dans le monde animal. En général, les animaux ont de la peine à établir des liens entre des événements temporellement et contextuellement distants et moduler leur comportement en conséquence (STEPHENS *et al.* 2002). Ainsi, on fait remarquer à WILKINSON que l'observation d'une corrélation entre des dons de sang et un retour de services ne nous dit encore rien sur les mécanismes sous-jacents à cette corrélation. En d'autres termes cette corrélation n'est pas une preuve suffisante pour postuler une situation de jeu itéré avec application de la stratégie *Donnant Donnant* (DE WAAL 2000 ; SILK 2003 ; HAMMERSTEIN 2003b). D'autres interprétations concurrentes doivent être envisagées. On pourrait par exemple imaginer que la réciprocité chez les vampires est basée sur des relations privilégiées : les individus pourraient être dotés d'une tendance à apprécier certains partenaires plus que d'autres sur la base de caractéristiques comme une odeur similaire à celle de leurs proches parents ou la production de sons typiques au groupe

dont ils font partie (HAMMERSTEIN 2003b ; LEHMANN & PERRIN 2002).[98]

On peut observer d'autres formes de réciprocité dans le monde animal, mais tout comme pour l'exemple des vampires, rien ne prouve qu'il s'agisse de formes d'altruisme réciproque. Un bon nombre d'observations faites sur les singes (capucins, macaques et babouins) montrent que le service de l'épouillage peut être échangé contre une série d'autres services : épouillage en retour, non-agression (un individu du bas de la hiérarchie qui épouille régulièrement un dignitaire du haut de la hiérarchie sera moins souvent agressé par ce dernier), accès privilégié aux ressources alimentaires, etc. Toutefois, sur le terrain, les facteurs qui déterminent la monnaie d'échange ainsi que les mécanismes sous-jacents à ces échanges sont extrêmement difficiles à saisir si bien qu'il n'est pas possible de montrer que toutes les conditions nécessaires à l'altruisme réciproque du modèle *Donnant Donnant* sont effectivement réalisées (BARRETT *et al.* 1999 ; MANSON *et al.* 2004).

En fin de compte, les exemples les plus convaincants d'altruisme réciproque (par exemple chez les singes ou certains ongulés) sont des formes d'épouillage mutuel en alternance rapide : l'effort est parcellisé (je t'épouille 2 minutes et tu m'épouilles 2 minutes en retour, etc.) ce qui permet de contrôler le partenaire de l'interaction ; cette succession rapide de dons et de retours de services facilite la réciprocité puisqu'elle permet un meilleur contrôle du partenaire de l'interaction et n'exige pas de capacités cognitives trop développées (comme celles qui permettent de lier des événement temporellement et contextuellement distants). Ces comportements d'épouillage mutuel en alternance rapide ont notamment été observés chez les impalas (CONNOR 1995) et les babouins (BARRETT *et al.* 1999).[99] Notons ici que l'intuition de G. ROBERTS et SHERRATT (1998) s'avère en partie juste : la réciprocité a plus de chances de fonctionner s'il y a alternance rapide entre les dons et les retours de services.

[98] Une autre interprétation serait d'expliquer le don de sang chez les vampires à l'aide de modèles qui ne font même pas entrer en jeu la réciprocité ! Rick RIOLO, Robert AXELROD et collègues (2001) proposent par exemple une explication de l'émergence d'actions coopératives (non liées à des retours de services) dirigées exclusivement vers des individus similaires à eux-mêmes par rapport à certains traits observables.

[99] Notons à ce propos que les modèles de théorie des jeux proposés par AXELROD puis G. ROBERTS et SHERRATT ne tiennent pas compte de ce facteur de réciprocité à court terme.

Bilan

Nous avons vu quelles étaient les conditions nécessaires au développement d'un comportement altruiste réciproque. Malheureusement, il semblerait qu'il soit rare dans le monde animal. Les symbioses entre nettoyeurs et poissons prédateurs ne remplissent pas toutes les conditions de l'altruisme réciproque puisque l'échange de service se fait de manière simultanée ; l'exemple de l'échange de sang chez les vampires n'est pas entièrement convaincant puisque WILKINSON n'a pas pu donner la preuve de la pratique d'un comportement punitif en cas de défection et que des interprétations concurrentes semblent plus pertinentes. Il reste les cas de l'épouillage mutuel mais il s'agit d'une forme très primaire d'altruisme réciproque puisqu'elle est basée sur un échange à alternance rapide de services du même type. Il semblerait que la forme subtile d'altruisme réciproque sur laquelle AXELROD met l'accent exige un ensemble de capacités cognitives développées (mémoire sélective, conceptualisation, catégorisation) que l'on trouve uniquement chez les êtres humains.

Quoi qu'il en soit, si l'on considère les gènes responsables du comportement de l'épouillage mutuel, on constate qu'ils peuvent uniquement se répandre au fil des générations dans le pool génétique d'une population si cette forme de réciprocité bénéficie en fin de compte aux individus eux-mêmes. Cette conclusion nous mène à la question suivante : Dans ce contexte de réciprocité, est-il encore légitime de parler d'altruisme ? Rappelons ici sa définition : un comportement est dit altruiste s'il a pour effet d'augmenter la *fitness* individuelle d'autrui aux dépens de la *fitness* de l'individu qui développe ce comportement. Or un comportement altruiste réciproque n'a justement pas, par définition, pour effet d'augmenter la *fitness* d'autrui aux dépens de sa propre *fitness* ; au contraire, dans des circonstances favorables, il a pour effet d'augmenter la *fitness* de l'individu qui développe ce comportement. En conséquence, je pense qu'il faut admettre que les cas d'altruisme réciproque n'entrent pas sous la définition de l'altruisme évolutionnaire. A ce propos, je dirais même qu'il s'agit d'un abus de terme ; dans ce contexte il vaudrait mieux utiliser des termes comme « réciprocité », « coopération » ou « investissement à long terme ».[100]

[100] Il convient de remarquer que lorsqu'elle est utilisée dans le cadre de la théorie des jeux, la notion d'altruisme évolutionnaire est passablement simplifiée. En effet, en théorie des jeux la *fitness* peut uniquement être calculée en fonction des coûts et gains (souvent compris en termes d'unités monétaires) pour les individus.

En conclusion, la théorie de la réciprocité porte à croire que beaucoup de comportements à première vue altruistes s'avèrent en fin de compte ne pas l'être puisqu'ils se soldent par un avantage individuel sur le long terme. Cette théorie apporte de l'eau au moulin de la controverse autour de l'altruisme évolutionnaire et fait nettement pencher la balance du côté des stratèges de l'égoïsme. A la section suivante, nous verrons que cette direction est encore renforcée par la théorie du signal coûteux.

2.2.3. La théorie du signal coûteux

Il existe une théorie similaire à celle de l'altruisme réciproque qui permet d'expliquer certaines actions « apparemment altruistes ». Il s'agit de la théorie du signal coûteux[101] selon laquelle un organisme peut investir de l'énergie pour produire un signal qui lui rapportera un avantage individuel en retour.

Un signal est un trait perceptible ou une action qui a évolué parce qu'il/elle indique la possession, par l'individu qui signale, d'un trait qui autrement resterait imperceptible ;[102] il peut par exemple signaler que l'individu est un bon partenaire de coopération (FRANK 1988) ou un bon parti pour la reproduction (ZAHAVI 1977 ; 2002 ; G. ROBERTS 1998). Un signal est un indicateur *fiable* d'un trait si sa présence est toujours ou presque toujours couplée avec la présence du trait.

La théorie du signal coûteux est toujours liée au problème de la tromperie. En effet, dès lors qu'un signal est pris au sérieux par les individus d'une population, on verra évoluer des signaux trompeurs qui permettent d'obtenir à moindre frais les gains liés à la production des signaux honnêtes. Dès lors, pour être fiable, un signal honnête se doit d'être coûteux ; plus précisément, un signal efficace est à la fois trop coûteux pour être produit de manière trompeuse et suffisamment avantageux pour pouvoir être produit tout court.

---

[101] Cette théorie est basée sur la théorie du handicap dont il était question à la section 1.1.3 (note 29).

[102] Il peut être intéressant de distinguer entre un signe (*cue*) et un signal (*signal*) (MAYNARD SMITH & HARPER 2003). Au contraire d'un signal, un signe n'a pas évolué parce qu'il a pour fonction de manifester l'existence d'un trait non observable. Par exemple, l'odeur d'un lapin est un signe (et non un signal) utilisable pour un prédateur lorsqu'il traque sa proie. Par contre, la longue queue d'un paon est un signal qui indique la bonne santé de l'individu (notons que cette longue queue, si elle est un signal pour les partenaires sexuels, est également un excellent signe pour les prédateurs du paon !).

Voici un exemple de signal coûteux tiré du monde animal. L'éthologue Amotz ZAHAVI (1977) a observé une espèce d'oiseaux (cratérope écaillé) organisée en système hiérarchique. Il a constaté que les individus du sommet de la hiérarchie produisent régulièrement des actions « apparemment altruistes » ; ils nourrissent leurs congénères ou effectuent des tours de guet pour avertir les autres de l'arrivée d'un prédateur au lieu de consacrer ce temps à se nourrir eux-mêmes. Les individus du fond de la hiérarchie qui tentent de produire ce genre d'actions se trouvent immédiatement brimés par leurs supérieurs. Il semblerait que dans ce cas, les actions altruistes exercent une fonction de publicité que seuls les individus les plus influents peuvent se permettre d'exercer. Ainsi, dans des conditions favorables, développer le signal coûteux de l'altruisme peut produire un avantage sélectif individuel (étant entendu qu'une bonne place dans la hiérarchie facilite l'accès à la reproduction).

Bilan

Tout comme la théorie de la réciprocité, la théorie du signal coûteux explique des comportements « apparemment altruistes » en termes d'avantages sélectifs, à la fois en faveur des gènes (comme c'est le cas dans la théorie de la sélection de parentèle) et des individus eux-mêmes ; dans des conditions favorables, le signal coûteux a pour effet d'augmenter la *fitness* de l'individu porteur du signal.

Pour ce qui est de la controverse autour de l'altruisme évolutionnaire, force est de constater que le camp des stratèges de l'égoïsme est bien fourni (HAMILTON, TRIVERS, MAYNARD SMITH, AXELROD, G. ROBERTS, HAMMERSTEIN, ZAHAVI, etc.). Reste à savoir si la théorie de la sélection de parentèle, combinée à celles de la réciprocité et du signal coûteux, permettent de rendre compte de tous les cas d'altruisme rencontrés dans le monde animal. Ce n'est pas l'avis de tout le monde, comme nous verrons à la section suivante.

2.2.4. La sélection génétique de groupe

La théorie de la sélection de groupe a déjà été thématisée par Charles DARWIN pour expliquer l'évolution des comportements altruistes.[103] Elle a connu ensuite une longue période faste avant d'être critiquée dès les années 1960 par les défenseurs du point de vue du gène. Sous le coup de ces critiques, elle s'est effondrée avant de renaître sous la plume d'auteurs contemporains. C'est ce chemin mouvementé que nous allons retracer ici. Mais avant cela, quelques remarques préliminaires s'imposent.

De manière générale, la théorie de la sélection de groupe tient compte du fait que les espèces se déploient en groupes et qu'au cours du processus de sélection naturelle, il y a survie de certains et disparition des autres. Lorsqu'elle est appliquée à la question de l'évolution des comportements altruistes, cette théorie fonctionne de la manière suivante : bien que sur le plan individuel l'altruisme soit désavantageux (en termes de *fitness*), il s'avère bénéfique sur le plan du groupe ; les individus altruistes fournissent un avantage sélectif au groupe auquel ils appartiennent et puisque les groupes composés d'altruistes ont plus de chances d'être sélectionnés que ceux qui n'en comptent pas, ils sélectionneront dans leur sillage leurs membres altruistes.

A première vue, l'analyse paraît simple mais elle s'accompagne en réalité d'un certain nombre de problèmes d'interprétation. En voici deux. Le premier concerne la façon de concevoir le groupe. Un gène ou un individu sont des entités bien pratiques ; il n'y a pas moyen de se tromper sur l'objet désigné. Par contre, un groupe est quelque chose de plus flou. Certains diront que c'est une tribu, une population, une bande ou un clan qui peut avoir une durée de vie bien supérieure aux individus qui le composent (DARWIN 2000/1871 ; WYNNE-EDWARDS 1986 ; LORENZ 1977/1963). D'autres diront qu'un groupe est constitué d'individus qui interagissent et influencent mutuellement leur *fitness* par rapport à un trait particulier (par exemple l'altruisme) ; le groupe disparaît (mais pas les individus qui le composent) lorsque les individus n'ont plus l'occasion d'interagir par le biais de ce trait. (D. WILSON 1975 ; SOBER & D. WILSON 2003/1998, pp. 92-98)

[103] Darwin a utilisé l'idée de sélection de groupe (*community selection*) pour rendre compte du comportement altruiste humain dans *La filiation de l'homme* (2000/1871).

Le second problème d'interprétation concerne la manière d'envisager les mécanismes au moyen desquels la sélection de groupe est susceptible d'opérer : tantôt il s'agit de conflits directs entre groupes (DARWIN 2000/1871) ; tantôt il s'agit d'une compétition indirecte par le biais de la croissance et de la division de groupes (quand un groupe devient trop grand, il se divise) (HALDANE 1932) ; tantôt il s'agit d'un mécanisme savant de dissolution, mélange et recomposition périodique des groupes (SOBER & D. WILSON 2003/1998 ; voir sections 2.2.4.iii et 2.2.4.iv).

Ainsi, lorsqu'il s'agit de juger de la pertinence de la théorie de la sélection de groupe, il ne faut jamais perdre de vue qu'il en existe différentes conceptions possibles, selon la manière dont on conçoit le groupe et les mécanismes sous-jacents à ce type de sélection.

i. Les premières théories de la sélection de groupe

C'est dans *La Filiation de l'homme,* que DARWIN avance l'hypothèse de la sélection de groupe pour résoudre le paradoxe de l'altruisme. On s'en souvient, le père de la théorie de l'évolution avait observé certains traits véritablement altruistes dans la nature qui semblaient se soustraire à la logique de sa théorie de sélection naturelle ; par exemple, il avait observé le fait que certaines abeilles meurent après avoir piqué un intrus approchant leur ruche. Or, selon sa théorie de la sélection naturelle, un tel trait ne peut pas avoir été sélectionné ; il aurait dû être éliminé au même titre que tous les autres traits nuisibles aux individus qui les possèdent. Pour échapper au paradoxe, DARWIN a imaginé une solution mettant en jeu le niveau du groupe : il commence par remarquer que beaucoup de comportements altruistes, bien que nuisibles aux individus qui les exercent, ne sont pas pour autant inutiles au niveau du groupe ; ensuite, il observe que les groupes, tout comme les individus, sont en compétition constante dans la nature ; il en déduit qu'un trait bénéfique pour un groupe peut logiquement avoir été sélectionné en dépit du coût qu'il engendre pour l'individu qui le porte.[104]

La théorie de la sélection de groupe était bien acceptée et fréquemment utilisée par les biologistes entre les années 1930 et 1960. Certains l'utilisaient parallèlement à la

[104] DARWIN ne fait pas usage du langage qui j'ai utilisé ici pour présenter sa théorie ; il parle de « communauté », « tribu » et « sélection tribale ». Notons également qu'avec cette théorie, il contredit certains de ses écrits antérieurs. Pour davantage de détails à ce sujet, voir GAYON 1998, chap. 2.

théorie de la sélection traditionnelle, privilégiant selon les cas celle qui leur semblait la plus apte à expliquer leur objet de recherche ; ils faisaient appel à la théorie traditionnelle de la sélection pour expliquer des traits comme les dents longues, ou la résistance à certaines maladies, alors qu'ils s'appuyaient sur la théorie de la sélection de groupe pour expliquer d'autres phénomènes, tels que l'ordre dans lequel les membres d'un groupe ont accès à la nourriture (WYNNE-EDWARDS 1962). Par exemple, Konrad LORENZ (1977/1963) avait noté qu'entre individus d'une même espèce, les animaux ont tendance à refuser le combat pour ménager leurs congénères. Selon lui, cette retenue était un trait sélectionné en raison du bénéfice qu'il apportait au groupe.

La popularité de la théorie de la sélection de groupe tient probablement au fait que, contrairement à leurs successeurs, ces biologistes réfléchissaient plutôt en termes qualitatifs que quantitatifs ; ils n'avaient pas coutume d'élaborer ou d'utiliser des systèmes mathématiques complexes pour étayer leurs théories (à l'exception notoire de Sewall WRIGHT 1945).

ii. La disgrâce de la sélection de groupe

Dans le courant des années 60, l'hypothèse de la sélection de groupe devint la cible d'attaques répétées. La cause de ce rejet massif est sans conteste l'émergence de la perspective du gène (section 1.1.2) ; selon ses défenseurs, les traits n'évoluent pas parce qu'ils aident le groupe, ni parce qu'ils augmentent le bénéfice individuel mais parce qu'ils favorisent la réplication des gènes qui induisent ces traits. Citons DAWKINS pour une critique récurrente contre la théorie de la sélection de groupe :

> « S'il existe un seul rebelle égoïste prêt à exploiter l'altruisme du reste du groupe, alors, par définition, ce sera lui qui aura le plus de chances de survie et d'avoir des enfants. Chacun de ses enfants aura tendance à hériter de cet égoïsme. Après plusieurs générations de cette sélection naturelle, le « groupe altruiste » sera dépassé par le nombre d'individus égoïstes et ne pourra plus se démarquer du groupe égoïste. » (DAWKINS 1996/1976, p. 25)[105]

[105] En réalité, cet argument n'exclut pas la possibilité théorique de la sélection de groupe. Même les défenseurs les plus acharnés de la perspective du gène admettent ce point (MAYNARD SMITH 1964 ; George WILLIAMS 1966). Ils soutiennent en revanche qu'elle apparaît extrêmement rarement dans la réalité et qu'elle est tout simplement inutile pour expliquer les phénomènes naturels. Ils ajoutent que la

Voilà une répétition de la formulation du paradoxe de l'altruisme. L'argument est de taille et les défenseurs de la sélection de groupe ne deviendront crédibles qu'à condition d'affiner leur théorie et de montrer qu'une sélection en défaveur de l'altruisme à l'intérieur de chaque groupe n'est pas suffisante pour compenser le mouvement inverse qui s'effectue au niveau de la sélection de groupe. Pour ce faire, il faudra qu'à l'image de leurs opposants, ils s'initient aux calculs de la théorie des jeux et se lancent dans la modélisation de situations d'interaction. Il faudra également qu'ils tiennent compte des théories de la sélection de parentèle et de la réciprocité ; théories d'autant plus « dangereuses » qu'elles ont largement contribué au discrédit de la sélection de groupe en parvenant à résoudre précisément les dilemmes qui avaient poussé DARWIN, LORENZ et d'autres à émettre l'hypothèse de la sélection de groupe. HAMILTON (1964) n'a-t-il pas donné une explication convaincante des comportements étonnants des abeilles, rendant du même coup superflue l'explication imaginée par DARWIN ? MAYNARD SMITH (1982) a fait de même avec le refus, souvent observé chez les animaux, de combattre contre des individus de la même espèce. Il a développé un jeu itératif dans lequel il s'agit d'obtenir une ressource et où deux stratégies s'opposent : une stratégie qui induit un comportement agressif jusqu'à la victoire ou la mort (appelée « faucon ») et une stratégie de retraite face à l'imminence d'un combat (appelée « colombe »). Il a pu montrer qu'une longue série de confrontations (organisées selon le modèle de la sélection naturelle) entre des individus arborant ces deux stratégies aboutit généralement à un équilibre évolutionnairement stable composé d'une majorité de colombes. Ainsi, il n'est pas nécessaire, comme le pensait LORENZ (1977/1963), de recourir au mécanisme de la sélection de groupe pour expliquer l'évolution de comportements non agressifs. Ce que Lorenz expliquait en termes de bien pour l'espèce, MAYNARD SMITH peut le traduire en termes d'avantage individuel : le refus de combattre de la colombe n'a pas évolué parce qu'il bénéficie au groupe, mais parce que les colombes elles-mêmes tirent un avantage à ne pas se battre jusqu'à la mort. Une fois de plus, ce qui était considéré comme altruiste devient égoïste.

---

perspective du gène est bien plus élégante et économique du fait qu'elle permet de focaliser l'attention sur un seul niveau de sélection.

iii. La théorie de la sélection à multiples niveaux : une réhabilitation de la théorie de la sélection de groupe

Il aura ensuite fallu attendre plusieurs décennies pour que la théorie de la sélection de groupe reprenne son envol. Le modèle de sélection de groupe qui sera présenté en détails dans cette section a été élaboré par David Sloan WILSON. Quoique déjà formulé dans les années 1970 (D. WILSON 1975), ce modèle a gagné en popularité avec la parution, en 1998, de *Unto Others*, un ouvrage écrit en collaboration avec le philosophe des sciences Elliott SOBER. Selon les auteurs de ce livre, non seulement la théorie de la sélection de groupe se défend, mais en plus il s'agit d'un excellent outil théorique pour expliquer la sélection de comportements altruistes.

Selon eux, c'est une erreur de calculer les avantages sélectifs uniquement au niveau des gènes car ce ne sont pas les seuls bénéficiaires ou victimes de la sélection naturelle ; les organismes individuels et les groupes le sont également. Ainsi, si l'on veut appréhender rigoureusement le phénomène de la sélection naturelle, il faut tenir compte de trois niveaux de sélection : celui du gène, celui de l'individu et celui du groupe.[106] Au fond, c'est une manière de réhabiliter la sélection de groupe tout en préservant les acquis obtenus par les penseurs des trente ou quarante dernières années.

Pour illustrer l'erreur commise par les défenseurs de la perspective unique du gène, SOBER et D. WILSON présentent le « paradoxe de Simpson » (qui, en réalité n'est pas un paradoxe). Il s'agit d'un modèle mathématique tiré d'une situation réelle.

Une université est accusée de discrimination sexuelle car la proportion d'hommes admis est de 14 %, alors que la proportion de femmes admises n'est que de 12 %. On mène une enquête, dont les résultats montrent que dans chaque faculté considérée indépendamment (mettons qu'il y en ait deux : Lettres et Électronique), la proportion de femmes reçues est plus élevée que la proportion d'hommes reçus (même si la proportion globale est effectivement plus élevée chez les hommes que chez les femmes).

[106] Notons que cette idée n'est pas neuve. En 1970 déjà George PRICE et Richard LEWONTIN défendaient la hiérarchie de sélection.

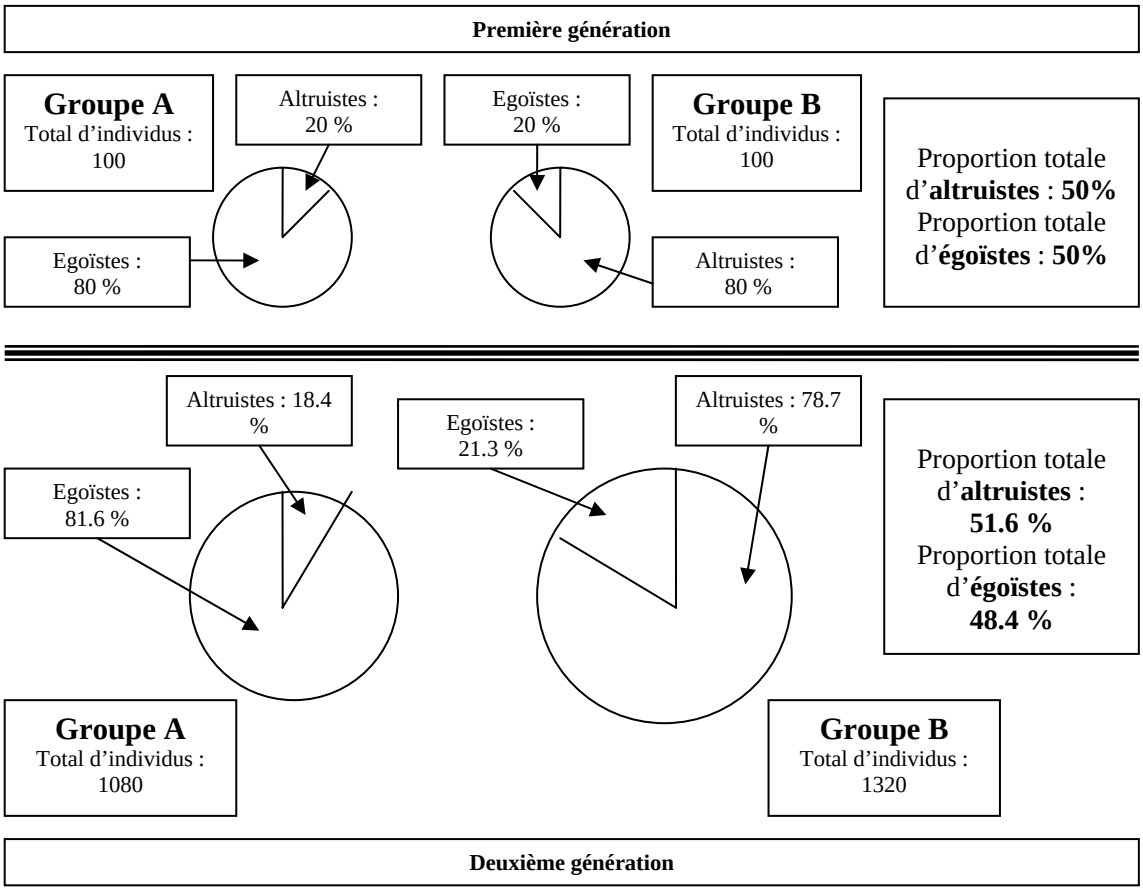
	Nombre de postulants de sexe masculin	Nombre de postulants de sexe féminin	Nombre de postulants admis de sexe masculin	Nombre de postulants admis de sexe féminin	Pourcentage de postulants admis de sexe masculin	Pourcentage de postulants admis de sexe féminin
Lettres	10	100	1	11	10%	11%
Electronique	40	10	6	2	15%	20%
Total	50	110	7	13	14%	12%

En examinant les chiffres du tableau (tiré de DE SOUSA 2004, pp. 109), on constate qu'il y a beaucoup plus de femmes qui postulent en Lettres qu'en Electronique (et inversement pour les hommes) et que la sélection est plus dure en Lettres qu'en Electronique. C'est en vertu de ces deux facteurs que la proportion de femmes admises est globalement inférieure à celle des hommes, alors même que dans chaque faculté considérée isolément, la proportion de femmes admises est supérieure à celle des hommes.

SOBER et D. WILSON proposent cet exemple contre-intuitif en guise d'analogie à ce qui se passe dans le phénomène de la sélection naturelle ; ils veulent montrer que la sélection est un phénomène complexe qui opère à plusieurs niveaux (par exemple à l'intérieur de chaque groupe et entre différents groupes). C'est pourquoi, en opposition à la perspective du gène, ils prennent le parti de réfléchir en termes de sélection à multiples niveaux, les trois niveaux étant celui des gènes, des individus et des groupes.

Voyons comment fonctionne cette sélection à multiples niveaux dans le cas de l'évolution de l'altruisme. Pour SOBER et D. WILSON, le phénomène décrit par le paradoxe de Simpson apporte un élément décisif à la résolution du paradoxe de l'altruisme : nous avons vu que ce dernier réside dans le fait que les altruistes sont condamnés à avoir une *fitness* perpétuellement inférieure à celle des non-altruistes avec lesquels ils cohabitent, si bien que le trait de l'altruisme semble être tragiquement voué à l'extinction au sein du groupe. Cependant, nous font remarquer les auteurs, plus la proportion d'altruistes dans un groupe est élevée, plus la *fitness* globale des individus qui le composent (altruistes comme non-altruistes) est élevée. Ainsi, il semble probable que l'altruisme ait pu évoluer en raison des bénéfices qu'il apporte à son groupe en dépit du fait que la proportion d'altruistes face aux non-altruistes diminue dans chaque

groupe considéré individuellement.[107] Le schéma ci-dessous (inspiré de SOBER et D. WILSON 2003/1998, p. 24) illustre cette idée.



Imaginons une population asexuée composée d'individus altruistes et d'individus non altruistes. Cette population est divisée en deux groupes de 100 membres chacun (ce qui correspond aux deux petits cercles du haut). Les altruistes composent 20 % du groupe A et 80% du groupe B. Ainsi, la proportion d'altruistes de la population globale est de 50%. A la génération suivante (qui correspond aux deux grands cercles du bas), la fréquence d'altruistes décroît dans chacun des deux groupes par rapport à la fréquence des non-altruistes (passage de 20% à 18.4% dans le groupe A et de 80% à 78.7% dans le groupe B). Par contre, le groupe B (composé d'une majorité d'altruistes) se porte mieux que le groupe A ; il est devenu plus grand (il compte 1'320 individus contre 1'080 pour

[107] Parallèlement, dans le paradoxe de Simpson, en dépit du fait que les hommes soient discriminés par la politique d'admission de chaque faculté considérée, ils finissent par être admis en plus grand nombre que les femmes. Cela tient au fait qu'ils ont majoritairement « parié » pour la faculté la moins sélective.

le groupe A). Résultat : la proportion d'altruistes de la population globale a légèrement augmenté : elle est passée de 50% à 51.6%.

Pour rendre ce modèle plus parlant, accompagnons-le d'une histoire : SOBER et D. WILSON (2003/1998) reprennent un exemple inventé par MAYNARD SMITH (1964), intitulé le « modèle des bottes de foin ».[108] Il envisage une race de souris vivant sur plusieurs générations en petits groupes entièrement isolés dans des bottes de foin. Parmi ces souris, certaines sont altruistes et d'autres non. Les altruistes perdent beaucoup de temps à épouiller leurs congénères (leur épargnant ainsi un bon nombre de maladies), comportement qui permet d'augmenter sensiblement la *fitness* moyenne du groupe. Au bout d'un certain nombre de générations,[109] les groupes se dissolvent (une fois par année, toutes les souris sortent en même temps des différentes bottes de foins) pour ensuite former de nouveaux groupes. Chaque groupe de départ est fondé par une seule souris femelle portante qui peut donner naissance à des souriceaux altruistes ou à des petits qui ne le sont pas. Ainsi, les individus au sein des groupes sont tous apparentés lors de la première génération ; les degrés de parenté s'estompent ensuite rapidement.[110] SOBER et D. WILSON pensent que si on attribue des valeurs réalistes aux fréquences de reproduction ainsi qu'aux coûts et gains liés aux comportements altruistes, la simulation de la sélection naturelle donnera les résultats suivants : A partir de la deuxième génération, la fréquence des altruistes baisse à l'intérieur de chaque groupe, sans pour autant causer la disparition de tous les altruistes avant la fin du cycle. Les groupes comprenant des altruistes se portent mieux (puisque leurs membres souffrent moins de maladie) que les groupes n'en comprenant pas. A la fin du cycle, lorsque les individus de tous les groupes se rassemblent en un même lieu, il est tout à fait réaliste d'imaginer que la proportion d'altruistes soit supérieure à celle du brassage de population antérieure au cycle ; c'est le cas si l'effet de sélection de groupe (favorable à l'altruisme) était suffisamment puissant pour compenser l'effet de sélection individuelle (défavorable aux

[108] MAYNARD SMITH, un défenseur de l'approche par le gène, a élaboré ce modèle précisément pour montrer que la sélection de groupe est uniquement possible dans des conditions extrêmement difficiles à réaliser si bien qu'il est très rare de la rencontrer dans le monde animal…

[109] Dans la réalité, les souris femelles atteignent leur maturité sexuelle à 1½ mois, la durée de gestation est d'environ 3 semaines et la période de reproduction s'étend sur toute l'année.

[110] Cette affirmation que le degré de parenté s'estompe est trompeuse ; si au début du cycle, il n'y a que deux parents, en fin de compte ce seront toujours les mêmes gènes qui seront recombinés dans la botte de foin. Cette critique sera reprise plus loin (p. 108).

altruistes) à l'intérieur de chaque groupe.

Il convient tout de même de remarquer que le calcul de la proportion globale d'altruistes et de non-altruistes se justifie uniquement si les groupes se désintègrent effectivement et que leurs populations respectives se mélangent dans le même bassin de population ; si ce n'est pas le cas, dans chaque groupe et au fil des générations, la sélection naturelle se chargera de la disparition des altruistes. Ainsi, le modèle proposé par SOBER et D. WILSON se différencie nettement de ceux des premiers défenseurs de la théorie de la sélection de groupe (DARWIN y compris) : il n'y a pas lieu ici de concevoir la sélection au sens d'un conflit direct entre différents groupes.

iv. De la théorie à la vie réelle : l'altruisme des mutants de la petite douve

Pour que la théorie de la sélection de groupe présentée par SOBER et D. WILSON puisse rendre compte de l'évolution de l'altruisme évolutionnaire, il faut que la formation et désintégration périodique des groupes qu'elle postule apparaisse effectivement dans la nature.

Comme exemple SOBER et D. WILSON (2003/1998) proposent celui du ver du cerveau (*brain worm*). Le *Dicrocoelium dendriticum* ou « petite douve » est un ver parasite qui commence et termine son cycle dans le foie d'un bovidé. Le cycle débute par une expulsion des œufs du ver parasite hors de la vache, via les excréments. Les œufs sont ensuite ingurgités par des escargots friands d'excréments. Pour faciliter la démonstration, SOBER et D. WILSON supposent que chaque escargot qui se nourrit des excréments du bovidé ingurgite invariablement 5 œufs. Ces 5 œufs (ou plutôt leurs habitants) ne se sépareront plus jusqu'à la fin du cycle ; ils constituent donc un groupe. A l'intérieur du mollusque, les vers éclosent et se décuplent par reproduction asexuée (ils créent des clones d'eux-mêmes). Ainsi, on obtient un groupe de 50 vers, qui est ensuite expulsé dans son entier par l'escargot sous forme d'une boule de mucus. Cette boule de mucus est ingurgitée par une fourmi dans l'estomac de laquelle les vers forment un kyste et attendent que leur hôte soit ingéré par un bovidé. Si le cas se présente, les vers terminent leur cycle dans le foie du bovidé où ils rencontrent des partenaires sexuels provenant d'autres groupes et se reproduisent de manière sexuée.

Dans la réalité, les deux premiers passages (du bovidé à l'escargot, et de l'escargot à la fourmi) ne sont pas problématiques et s'effectuent avec un pourcentage

relativement élevé, étant donné que les escargots se nourrissent copieusement des excréments des bovidés et que les fourmis apprécient le mucus des escargots. Cependant, le troisième et dernier passage est plus difficile et apparemment plus rare, car cette fois-ci, les fourmis ne font pas partie de l'alimentation de base des bovidés et se réfugient dans leur fourmilière à l'approche du crépuscule, lorsque les bovidés commencent à se nourrir avec le plus d'ardeur. Ainsi, sauf quelques heureuses exceptions, de nombreux vers voient leur cycle interrompu dans l'estomac d'une fourmi et meurent sans descendance. Cependant, et c'est ce qui intéresse SOBER et D. WILSON, il existe un type spécifique de ver qu'ils appellent « ver du cerveau » (*brain worm*), issu d'une mutation, qui donne un véritable coup de pouce au passage de la fourmi au bovidé. Le comportement du ver du cerveau ne diffère pas de celui d'un ver normal jusqu'à ce qu'il soit ingurgité par une fourmi. Admettons qu'au départ, l'un des cinq oeufs était celui d'un ver mutant (A) entouré de quatre oeufs de vers normaux (S), on obtient dans l'estomac de la fourmi, par le miracle de la reproduction asexuée, un groupe G composé de 40 individus S pour 10 individus A. Et c'est dans l'estomac de la fourmi que le comportement des individus S et A se différencient : les S forment simplement un kyste dans l'estomac de la fourmi et attendent patiemment que leur hôte soit mangé par un bovidé. Parmi les A il en va autrement ; l'un des vers du cerveau, au lieu de former son kyste dans l'estomac de la fourmi et d'attendre, se dirige en direction du cerveau pour y former un kyste à un endroit bien précis, ce qui a pour conséquence de modifier le comportement de l'insecte ; en effet, lorsque la température baisse à l'approche du soir, la fourmi grimpe en haut d'un brin d'herbe (au lieu de se diriger vers son abri) et s'y accroche fermement à l'aide de ses mandibules ; ce comportement singulier a pour effet d'augmenter les chances que la fourmi se fasse ingurgiter par un bovidé. Toutefois, le ver à l'origine de cette manœuvre meurt sans avoir eu la possibilité de se reproduire, alors que les autres membres du groupe bénéficient du fruit de son sacrifice ; c'est en cela que l'on peut le considérer comme altruiste.

Cet exemple présente tous les ingrédients pour l'évolution de l'altruisme par la sélection de groupe ; on dispose de plusieurs groupes qui se forment à partir d'un seul bassin de population ; ces groupes durent le temps d'un cycle déterminé avant de se dissoudre à nouveau dans le bassin de population. Les groupes sont différents en matière de représentation d'altruisme ; certains contiennent des individus altruistes et d'autres pas. Enfin, il existe une relation directe entre l'existence d'altruistes et la *fitness*

globale des individus du groupe ; les groupes avec altruistes ont plus de chance d'achever leur cycle.

Demandons nous maintenant si le trait A (c'est-à-dire le comportement du ver du cerveau) peut être sélectionné au fil de l'évolution. Pour comprendre ce point, il est utile de présenter la manière dont SOBER et D. WILSON définissent la notion de *fitness*. Ce qu'ils appellent la « *fitness* relative » opère à deux niveaux: celui des individus au sein d'un groupe et celui des groupes. Dans le cas de l'exemple du ver du cerveau, la *fitness* relative des vers A diminue au sein du groupe, puisqu'en raison du ver qui se sacrifie, ils passent d'une population de 10 à 9 ; ainsi, la proportion des altruistes au sein du groupe chute de 20% à 18.3%. En ce qui concerne la *fitness* relative entre les groupes, il faut tout d'abord souligner que le ver du cerveau n'est pas absolument indispensable à l'accomplissement du cycle du parasite, puisque certaines fourmis sont accidentellement avalées par des bovidés même si elles ne s'accrochent pas au sommet des brins d'herbe. Par contre, les groupes qui contiennent des altruistes sont avantagés par rapport aux groupes qui en sont dépourvus, puisque l'action d'un des vers altruistes du groupe favorise le retour à la vache. Ainsi, si on admet qu'un groupe entièrement composé de vers S possède une *fitness* égale à F1, on peut dire qu'un groupe composé de vers altruistes possède une *fitness* supérieure à F1. Ensuite, il faut mettre en équation les deux *fitness* relatives (F du groupe et F de l'individu) pour déterminer en faveur de qui opère la sélection. Selon SOBER et D. WILSON, l'altruisme peut être sélectionné à condition qu'une augmentation (même minime) du taux de probabilité d'être ingéré par une vache compense le comportement « suicidaire » du ver du cerveau ;[111] dans ce cas, l'avantage au niveau du groupe l'emporte sur le déficit au niveau individuel. En bref, les groupes mixtes ont une *fitness* supérieure aux groupes constitués entièrement de non-altruistes, mais il faut que cette différence soit suffisante pour compenser la diminution en proportion des altruistes à l'intérieur des groupes qui terminent le cycle.[112]

[111] Notons que dans l'exemple du ver du cerveau, la différence entre les *fitness* des différents individus n'est pas due à un taux de reproduction différentiel puisque l'on postule que tous les parasites ont le même nombre de descendants (en l'occurrence, chacun produit 9 clones). Elle est uniquement due à la survie ou à la mort des individus au cours du cycle (soit une mort altruiste, soit une mort parce que le cycle ne s'est pas achevé).

[112] « Between-group selection favors the evolution of altruism; within-group selection favors the evolution of selfishness. These two processes oppose each other. If altruism manages to evolve, this

Malheureusement pour SOBER et D. WILSON, cet exemple du ver du cerveau, quoique divertissant, donne une forte impression d'invraisemblance. Il est vrai que la petite douve (*Dicrocoelium dendriticum*) existe et se comporte à peu près comme l'indiquent les auteurs. Mais les détails concrets des étapes par lesquelles passe ce parasite (le nombre d'œufs ingurgités par les escargots, la quantité de mucus ingérée par les fourmis, etc.) sont très difficilement observables. De plus, l'idée des mutants altruistes qui parasiteraient le cerveau des fourmis n'est qu'une hypothèse. En réalité, SOBER et D. WILSON sont forcés d'admettre que les détails conceptuellement pertinents pour leur modèle sont le fruit de simples suppositions.[113]

v. Le caractère englobant de la théorie de la sélection à multiples niveaux

Peut-on trouver des exemples plus convaincants de sélection de groupe que celui du ver du cerveau ? D'après SOBER et D. WILSON, il est aisé d'en trouver car, selon eux, la sélection à multiples niveaux englobe tous les cas de sélection de parentèle ;[114] au fond, affirment-ils, les individus parents qui s'entraident peuvent être considérés comme formant un groupe. Dans le cas des soins parentaux par exemple, les petites unités parents-enfants forment des groupes qui se dissolvent dès que les petits deviennent indépendants. Ensuite, il suffit de faire fonctionner la théorie de la sélection à multiples niveaux pour expliquer l'évolution des comportements altruistes.

La conception de la sélection et de l'évolution de l'altruisme proposée ici par SOBER et D. WILSON est extrêmement hasardeuse, puisqu'elle implique que l'on jette aux orties la fameuse théorie de la sélection de parentèle de HAMILTON. SOBER et D. WILSON cherchent à évincer cette théorie en portant l'attention non plus exclusivement sur les gènes mais également sur les individus et les groupes : selon eux, lorsqu'un

---

indicates that the group-selection process has been strong enough to overwhelm the force pushing in the opposite direction. » (SOBER & D. WILSON 2003/1998, p. 33)

[113] « …The conceptually relevant details have only been guessed ». (SOBER & D. WILSON 2003/1998, p. 30)

[114] Pour corroborer leur théorie, les auteurs proposent d'autres exemples de sélection de groupe comme celui de la répartition inégale des sexes chez une espèce d'araignée (2003/1998, pp. 38-43) ou celui de la rapidité de transmission des virus (2003/1998, pp. 43-46) ; mais il n'est pas forcément évident que l'on puisse parler d'altruisme évolutionnaire dans ces cas, car il n'y a pas vraiment d'interaction entre des individus altruistes et d'autres individus qui bénéficient directement du comportement des premiers.

individu altruiste aide un individu parent égoïste, on ne peut nier que du point de vue individuel, le premier sacrifie une partie de sa *fitness* (comprise ici au sens classique) au profit du second ; en d'autres termes, même à l'intérieur d'un groupe d'individus apparentés, les altruistes ont une *fitness* plus faible que les égoïstes.[115] Remarquons en passant qu'en donnant cette explication, SOBER et D. WILSON remettent au goût du jour la vieille théorie de la *fitness* classique où l'on tient uniquement compte de la viabilité et fécondité des individus (sans considérer l'alchimie génétique sous-jacente). Puis ils proposent leur théorie de la sélection de groupe en affirmant que si l'altruisme peut évoluer, c'est parce qu'il est favorable au groupe d'individus apparentés.[116] En d'autres termes, les auteurs soutiennent que l'altruisme ne peut en aucun cas être adapté du point de vue de la sélection individuelle; par contre, il peut l'être du point de vue de la sélection de groupe.

SOBER et D. WILSON ajoutent que la force de leur théorie de la sélection à multiples niveaux réside en ce qu'elle permet non seulement d'englober des comportements altruistes envers des proches parents, mais également envers des individus non parents (ce dont la sélection de parentèle est incapable). Les auteurs insistent beaucoup sur l'idée que l'altruisme n'est pas forcément lié à la parenté.

vi. Quelques doutes sur la théorie de la sélection à multiples niveaux

L'approche de SOBER et D. WILSON a beaucoup séduit les philosophes des sciences. Il est vrai que l'idée de sélection à multiples niveaux est très élégante. Il s'agit d'une nouvelle manière d'approcher la question de l'évolution de l'altruisme sans pour autant renoncer aux anciens outils conceptuels (*fitness*, gène, phénotype, mutation, sélection individuelle, sélection génétique, sélection de groupe…) ; la nouveauté réside

[115] Selon SOBER et D. WILSON, c'est une grande erreur de la part de la théorie de la sélection de parentèle, que de ne pas considérer le caractère altruiste des individus dits « altruistes » ; la sélection naturelle opère *toujours* contre les individus altruistes à l'intérieur des groupes mixtes (composés d'altruistes et d'égoïstes). « Inclusive fitness view (…) loses sight of the fact that natural selection operates against altruism (apparent or otherwise) in all mixed groups. » (2003/1998, p. 67).

[116] « When an altruistic individual helps a related individual who is selfish, the donor still has a lower fitness than the recipient. The fact that they are related does not cancel this fundamental fact. Within a group of relatives, altruists are less fit than selfish individuals. It is only because of selection among groups that altruism can evolve. » (D. WILSON & SOBER 2002, pp. 192-193)

dans la prise en compte des différents niveaux auxquels opère la sélection naturelle (niveaux des gènes, des individus et des groupes). Cette approche permet de mettre en évidence les conflits d'intérêts d'un niveau de la hiérarchie biologique à l'autre : ce qui est bon pour un gène peut être mauvais pour l'individu ; ce qui est positif pour un individu risque d'être nocif pour le groupe, etc. Elle permet également de comprendre que les bénéficiaires (ou les défavorisés) de la sélection naturelle ne sont pas uniquement les gènes ; les individus et les groupes peuvent également l'être (même si en un sens ils sont les supports des gènes).[117] Enfin, elle redonne à la fois à la *fitness* classique et à la sélection de groupe leurs lettres de noblesse.

Toutefois, malgré son élégance, cette théorie est loin de faire l'unanimité chez les biologistes de l'évolution et les critiques ne manquent pas.

Notons d'abord que pour pouvoir prétendre englober la théorie de la sélection de parentèle, SOBER et D. WILSON sont forcés d'adopter une notion de groupe flexible au point où l'on peut admettre que deux individus suffisent à former un groupe. Or on peut se demander s'il est encore utile de parler de groupe dans ces cas là (MAYNARD SMITH 1998).

D'autre part, le modèle de SOBER et D. WILSON ne fonctionne qu'en cas de fréquente dissolution et reconstitution des groupes ; il faut que les cycles se succèdent assez rapidement pour contrer l'effet de sélection à l'intérieur des groupes. Or, pour peu que les groupes soient composés de plus de deux individus non parents, la rareté des exemples pertinents se fait cruellement sentir.

Ensuite, leur théorie de la *fitness* relative implique que l'on admette l'existence d'une force sélective au niveau des groupes si bien que l'on peut parler de *fitness* de groupe. Mais que faut-il comprendre par-là ? Il est clair qu'un groupe ne peut pas avoir de descendance ; il n'y a pas de groupe maman qui engendre un groupe enfant. Or on se demande comment il est possible de parler de *fitness* si on ne peut pas calculer de

[117] Il est vrai qu'à priori, faire des gènes le seul objet de sélection paraît exagéré ; en principe, rien n'empêche de considérer les individus, voire les groupes d'individus comme des objets de sélection. Toutefois, il faut ajouter au crédit des défenseurs de la perspective du couple gène/phénotype, que si l'on observe l'évolution sur le long terme, ce sont les traits et les gènes responsables de ces traits qui sont sélectionnés ; le niveau de sélection des gènes est le plus pertinent car il porte sur des unités non recombinantes (c'est-à-dire qui ne se modifient pas d'une génération à l'autre).

descendance (GILDENHUYS 2003).[118] Cette critique est d'autant plus parlante lorsque l'on sait qu'il n'existe aucune valeur mathématique dans le modèle proposé par SOBER et D. WILSON représentant ce qu'ils appellent la « *fitness* de groupe ».[119]

D'autre part, déjà dans un article de 1976, MAYNARD SMITH critiquait la première formulation de la théorie de la sélection de groupe proposée par D. WILSON (1975). Selon MAYNARD SMITH, si on fait abstraction de la parure théorique et que l'on se concentre sur les calculs utilisés par D. WILSON, on retrouve la règle de Hamilton ! Ainsi, du point de vue mathématique, les deux approches se valent. Dès lors, la question est de savoir si, compte tenu de l'équivalence mathématique, il est encore utile de dire avec SOBER et D. WILSON, que la sélection agit à plusieurs niveaux, c'est-à-dire qu'elle exerce une *force* au niveau des individus et au niveau des groupes. MAYNARD SMITH (1976) et bien d'autres biologistes dans son sillage ne le pensent pas (GILDENHUYS 2003, Nicolas PERRIN et Laurent KELLER, communications personnelles). Pour eux, ce que propose D. WILSON n'est rien de plus qu'un modèle d'assortiment non aléatoire de gènes qui induisent des comportements altruistes ; c'est-à-dire un modèle où les individus altruistes ne dispensent pas leurs bienfaits au hasard mais de préférence (où plutôt avec une plus grande probabilité) en faveur d'autres individus altruistes. Cela n'empêche pas que l'on puisse parler de dynamique de groupe ; mais alors il faut réfléchir en termes de voisinage social et voisinage de compétition tels qu'ils ont été définis à la section 2.2.1.i (p. 60).

Que penser maintenant de l'affirmation de SOBER et D. WILSON selon laquelle leur théorie est plus englobante que celle de la sélection de parentèle de HAMILTON ? En

[118] « The groups do not produce offspring, or at least if they did, they would produce offspring groups rather than individual organisms. But the latter possibility is explicitly denied by Sober and Wilson: it is of crucial importance to the operation of the model that the members of any one set of subgroups formed by periodic subdivision of the global population recombine after interaction into a global population from which new subgroups are formed with a different assortment of members. The subgroups in the model do not autonomously or independently go on to produce the next set of subgroups. Thus, the analogy with Darwinian selection is misplaced since, according to Darwin, individuals that are more fit go on to produce other individuals that are more fit, while in Sober and Wilson's model, individuals subgroups that are more fit do not go on to produce individual subgroups that are more fit. » (GILDENHUYS 2003, pp. 32-33)

[119] « Despite their repeated use of 'group fitness' to describe what is going on in their model, there is no value in their mathematical analysis for the term. » (GILDENHUYS 2003, pp. 30-31)

un sens ils ont raison. Il est théoriquement possible qu'un comportement altruiste se répande dans une population d'individus non parents. HAMILTON l'avait pourtant déjà fait remarquer en 1970 ! Souvenons-nous de la notion de « *r* élargi » (section 2.2.1.i), le coefficient d'apparentement qui n'est pas forcément dépendant des liens parentaux (à cette occasion, j'avais utilisé le terme de « coefficient de relation génétique » ; p. 63). De manière plus explicite, dans un article de 1975, HAMILTON distingue la théorie de la *fitness* inclusive de la théorie de la sélection de parentèle, la première étant plus englobante. Il précise qu'il est théoriquement possible que l'altruisme soit sélectionné dans un milieu d'individus non parents. Selon lui, on peut imaginer deux autres cas où les altruistes ont de grandes chances de sacrifier une part de leur *fitness* au profit d'autres altruistes : soit les individus sont capables de se reconnaître entre eux et se sacrifient de préférence pour d'autres altruistes[120] ; soit les gènes qui incitent au comportement altruiste sont également responsables d'une préférence pour un type d'habitat particulier (dans ce contexte, on parle d'effets pléiotropique des gènes), si bien qu'en somme les individus altruistes se côtoient quotidiennement (ce qui augmente grandement les chances pour que les bénéficiaires des actions altruistes soient eux-mêmes porteurs des gènes pour l'altruisme).[121] La *fitness* inclusive permet de rendre compte à la fois des cas d'altruisme entre individus parents et les cas d'altruisme entre individus non parents.

Cela dit, il semblerait que l'apparentement génétique soit, de loin, le meilleur moyen de produire des adaptations altruistes (MAYNARD SMITH 1976 ; OKASHA 2002 ; DAWKINS 1979, p. 188). Considérons le cas où un gène a pour effet pléiotropique à la fois un comportement altruiste et un trait observable (par exemple une barbe verte). Dans ce cas, les individus altruistes ont la possibilité de diriger leurs bienfaits vers les autres altruistes ; si c'est le cas, l'altruisme peut évoluer dans un milieu d'individus non parents. Toutefois, comme le fait remarquer Samir OKASHA, ce scénario est confronté

[120] Cela est possible si tous les individus altruistes possèdent également un trait observable distinctif ; il s'agit de ce que DAWKINS appellerait l'« effet barbe verte » (voir section 2.2.1.i, p. 63).

[121] « Kinship should be considered just one way of getting positive regression of genotype in the recipient, and that it is this positive regression that is vitally necessary for altruism. Thus the inclusive-fitness concept is more general than 'kin-selection' (…). In the assortive-settling model it obviously makes no difference if altruists settle with altruists because they are related (perhaps never having parted from them) or because they recognize fellow altruists as such, or settle together because of some pleiotropic effect of the gene on habitat preference. » (HAMILTON 1975, pp. 140-141)

au risque de la tricherie car si, au fil de l'évolution, un individu non altruiste à barbe verte apparaît, il profitera des bienfaits des altruistes sans rien donner en retour, prospèrera, aura beaucoup d'enfants… et on imagine le reste de l'histoire. Pour éviter la tricherie, il faudra ou bien que les individus altruistes deviennent suffisamment intelligents pour détecter la tricherie, ou bien que le trait observable lié à l'altruisme soit difficile ou coûteux à imiter, les deux options étant très coûteuses. Or la sélection de parentèle ne souffre pas du problème de la tricherie (puisqu'un individu vivant aux dépens d'un de ses proches parents, empêchera la transmission d'une bonne partie de ses propres gènes) ; c'est donc le mécanisme le plus robuste pour l'évolution de l'altruisme (OKASHA 2002, p. 146).

Considérons maintenant le cas où un gène a pour effet pléiotropique à la fois un comportement altruiste et une préférence d'habitat. Dans une telle situation, des individus altruistes s'entraident de fait puisqu'ils vivent côte à côte si bien que le facteur de la parenté n'est pas nécessaire à l'évolution de l'altruisme. Toutefois, selon OKASHA, il est assez improbable que ce scénario se produise car il est extrêmement fragile. En effet, si les individus altruistes ne sont pas parents, cela signifie qu'ils ne partagent pas les mêmes gènes, sauf celui de l'altruisme qui leur fait également préférer l'habitat dans lequel ils évoluent. Ainsi, le gène qui induit un comportement altruiste ne profite qu'à lui-même (ou plus précisément à la production de copies de lui-même) et travaille au détriment de tous les autres gènes portés par l'individu altruiste.[122] Dans ces conditions, s'il y a apparition d'un gène qui occasionne les mêmes bénéfices que celui de l'altruisme (en l'occurrence la préférence d'habitat) sans être aussi coûteux du point de vue individuel, le gène responsable de l'altruisme sera rapidement voué à l'extinction via la sélection naturelle.[123] Par contre, il en va autrement si les bénéficiaires de l'altruisme sont des proches parents puisqu'ils partagent une bonne partie de leurs

[122] « From the point of view of all the other genes in an altruistic individual, the altruistic behavior is a waste. (…) While the recipients of altruism are also likely to carry copies of the altruistic gene, with respect to all other loci donor and recipient are no more likely than average to share genes. » (OKASHA 2002, p. 142)

[123] « From the point of view of all the genes in an altruistic organism except the altruistic gene itself, helping unrelated altruists is wasteful, so any way of preventing the waste without foregoing the associated benefit will be favored by natural selection. » (OKASHA 2002, p. 143)

gènes ; dans ce cas, il n'y a pas de pression sélective à d'autres loci du génome de l'individu porteur du gène responsable de l'altruisme.[124]

En définitive, il semble très improbable que l'altruisme évolutionnaire évolue dans un contexte autre que celui de sélection de parentèle. Mais que dire alors des exemples proposés posés par SOBER et D. WILSON (le ver du cerveau et les souris dans les bottes de foin) ? En observant attentivement l'exemple de la petite douve (section 2.2.4.iv), on constate que le sacrifice du ver du cerveau profite fortement à ses clones (puisqu'il y a eu reproduction asexuée dans l'estomac de l'escargot), c'est-à-dire à des individus qui partagent 100% de leurs gènes avec l'altruiste ![125] Cette réalité discrédite évidemment la connotation altruiste que lui confèrent les auteurs. En réalité, il suffit de recourir à la bonne vieille théorie de la sélection de parentèle de HAMILTON pour expliquer l'évolution du comportement du ver du cerveau. En fin de compte, la théorie de la sélection de groupe s'avère plus fourvoyante qu'utile.

Reprenons pour terminer le modèle des souris des bottes de foin (section 2.2.4.iii, p. 98). SOBER et D. WILSON affirment qu'après quelques générations, les souris ne sont plus liées par un coefficient d'apparentement significatif. Or sans parenté, la théorie de la *fitness* inclusive de HAMILTON n'est plus pertinente pour expliquer le comportement des souris altruistes. Toutefois, cette affirmation peut légitimement être mise en doute. En effet, dans le modèle en question, le groupe ne compte qu'une mère et ses petits ; il n'y aura donc que trois allèles de chaque gène (deux provenant de la mère et un provenant du père absent) qui se recombineront au fil des générations d'accouplements entre la mère et ses petits et les petits entre eux. Comment, dans ces conditions, peut-on

[124] « Matters are very different if the recipients of altruism are genetic relatives of the donor. (…) Donor and recipient are relatives, so have the same degree of relatedness at every locus in the genome. This means that the altruistic behaviour benefits all genes in the genome equally, not just the gene that codes for the altruism. Therefore, a modifier gene which suppresses the altruistic behaviour will undermine its own replication prospects. This is because an individual possessing this modifier ceases to behave altruistically towards kin – thereby foregoing the opportunity of assisting other individuals who have a greater than average chance of carrying a copy of the modifier gene themselves. » (OKASHA 2002, p. 143)

[125] SOBER et D. WILSON semblent pourtant en avoir conscience puisqu'ils admettent que grâce à la survie des autres individus altruistes, le degré du sacrifice n'est pas aussi extrême que ce qu'il y paraît au premier abord. « It is interesting, however, that the degree of sacrifice is much less than it first appeared, because the extreme sacrifice of the brain worm is diluted by the survival of the A types that did not become the brain worm. » (2003/1998, pp. 27-28)

légitimement supprimer le facteur de la parenté après quelques générations afin de balayer toute interprétation en termes d'avantages génétiques ?

Bilan

Quelles conclusions peut-on tirer au terme de cette analyse de la sélection de groupe ? En ce qui concerne la controverse autour de l'altruisme évolutionnaire, nous pouvons placer les théories de sélection de groupe du côté des romantiques. Les défenseurs de ces théories (LORENZ, D. WILSON, etc.) pensent que l'altruisme évolutionnaire s'avère effectivement sélectivement désavantageux. Ce n'est qu'en observant la dynamique des groupes que l'on comprend comment les comportements altruistes peuvent évoluer.

Nous avons vu qu'il existe différentes sortes de théories de sélection de groupe et avons pris le temps d'analyser l'une d'entre elles dans le détail : la théorie de la sélection à multiples niveaux proposée par SOBER et D. WILSON. Ces auteurs insistent beaucoup sur le fait que, du point de vue individuel, lorsqu'un individu se comporte de manière altruiste en faveur d'un autre individu (qu'il soit parent ou non), il sacrifie une partie de sa *fitness* classique au profit du second. De ce fait, l'altruisme ne peut en aucun cas être adapté du point de vue de la sélection individuelle. Voilà une manière de dire qu'il existe des cas authentiques d'altruisme évolutionnaire dans le monde animal et de se placer du côté des romantiques dans la controverse autour de l'altruisme évolutionnaire.

Pour expliquer l'évolution de l'altruisme, SOBER et D. WILSON font appel à la force de sélection de groupe, censée compenser les effets négatifs de la sélection individuelle. Nous avons vu que le modèle mathématique de SOBER et D. WILSON n'est pas faux ; par contre, l'explication théorique en termes de force de sélection de groupe qui lui est associée n'est pas aussi convaincante qu'il y paraît au premier abord. Finalement, on peut se demander s'il ne vaut pas mieux revenir à la bonne vieille théorie de la *fitness* inclusive proposée par HAMILTON, un membre du camp des stratèges de l'égoïsme.[126] De nos jours, ce débat est encore ouvert.

[126] David QUELLER exprime bien ce point de vue : « It has become clear in recent years that the same behaviors can also often be understood as a form of group selection – not the old group selection of Wynne-Edwards, but nevertheless a method that involves partitioning of selection into within-group and

2.2.5. Bilan sur la controverse

Au terme de notre analyse de la théorie de la sélection de groupe, faut-il penser que les stratèges de l'égoïsme ont gagné la controverse ? La réponse me semble vaine car en réalité, cette controverse n'en est pas vraiment une. A quelques exceptions près (ZAHAVI 2002/2000), les stratèges de l'égoïsme ne cherchent pas à nier l'existence de l'altruisme calculé en termes de *fitness* classique. Leur projet est d'expliquer en termes d'avantages évolutionnaires pourquoi des comportements apparemment altruistes ont pu être sélectionnés. Nous avons vu que s'il y a réciprocité ou signal coûteux, ces comportements n'ont que l'apparence de l'altruisme ; en fin de compte, du point de vue individuel, ils s'avèrent avantageux sur le long terme. Par contre, les comportements dont on peut expliquer l'évolution par la sélection de parentèle sont effectivement altruistes du point de vue individuel même si, en termes de *fitness* inclusive, ils s'expliquent par un avantage génétique.

Ainsi, lorsque SOBER et D. WILSON prétendent sauver la cause de l'altruisme dans leur livre *Unto Others*, ce n'est que rhétorique puisque dans les faits, ils claironnent ce que presque personne ne nie ; l'existence de l'altruisme défini en termes de *fitness* classique. De plus, malgré les apparences, SOBER et D. WILSON procèdent exactement de la même manière que les stratèges de l'égoïsme pour expliquer l'évolution des comportements altruistes : ils calculent la transmission génétique.

Au fond, je pense que si l'on comprend le langage métaphorique des stratèges de l'égoïsme, il n'y a plus lieu d'être choqué par leur soi-disant machiavélisme. Ils montrent simplement que, en arrière-fond des comportements altruistes, se dissimule un avantage au niveau génétique. Il est vrai que pour expliciter cette dynamique il leur arrive de parler des gènes comme d'entités égoïstes (y compris les gènes qui induisent des comportements altruistes) dont le seul but est de disséminer un maximum de copies d'eux-mêmes aux générations suivantes. Mais tout cela n'est que métaphore puisqu'il est évident que les gènes ne sont ni doués d'intention, ni même de pensée ; ils se

---

between-group components (see Sober and Wilson, 1998). But the fact remains that almost no one uses these methods much to think about and solve interesting problems. Each of the two methods can dissect social evolution into component parts, but where inclusive fitness divides nature neatly at the joints, other methods seem to hack clumsily through the long bones. » (QUELLER 2001, p. 263)

répliquent avec plus ou moins de succès, voilà tout. Ainsi, lorsque les stratèges de l'égoïsme utilisent le terme de « gène égoïste », ils veulent indiquer que la sélection naturelle ne permettra qu'aux gènes les mieux adaptés de se répliquer et se propager dans une population. D'autre part, lorsqu'il leur arrive d'utiliser le terme de « gène altruiste », il s'agit à nouveau d'un abus de langage ; c'est un raccourci pour signifier « gène qui induit un comportement altruiste ». En conséquence on pourrait même, sans se contredire, dire d'un gène qu'il est à la fois altruiste et égoïste.[127]

Si les stratèges de l'égoïsme n'avaient pas fait un tel usage du terme « égoïsme » et s'étaient contentés d'attribuer le qualificatif « altruiste » aux comportements et aux individus, il est fort probable que la controverse autour de l'altruisme évolutionnaire n'aurait pas fait couler tant d'encre.[128]

---

[127] Notons que le terme « gène égoïste » est parfois utilisé dans le sens d'opposé au gène altruiste : un gène qui induit un comportement non altruiste (dans ces cas, le contexte est généralement assez explicite).

[128] Considérons quelques exemples de mécompréhension.

Certains opposants aux stratèges de l'égoïsme, lorsqu'ils entendent dire que les comportements altruistes envers des individus parents ont pu évoluer parce qu'ils favorisent l'« intérêt des gènes » qui induisent ces comportements (au sens où ils augmentent leurs chances de réplication), croient entendre que les comportements en question sont le résultat d'un calcul d'intérêt égoïste. Or il est évident qu'aucun stratège de l'égoïsme ne serait disposé à soutenir cette dernière affirmation ! Comme exemple de ce type de critiques absurdes, citons l'anthropologue Marshall SAHLINS : « Il importe de noter, au passage, que les problèmes épistémologiques, que pose l'absence d'un support linguistique pour le calcul des coefficients de liaison r [coefficient d'apparentement], indiquent une carence grave de la théorie de sélection de parenté. Extrêmement peu de langues, de par le monde, connaissent les fractions : elles existent en indoeuropéen, et chez les civilisations archaïques de l'Orient, proche et extrême, mais font généralement défaut aux populations dites primitives. Les systèmes numériques dont disposent les chasseurs-collecteurs ne vont généralement pas au-delà de un, deux, trois. Quant à savoir comment des animaux s'y prennent pour déterminer que r (ego, cousins au premier degré) = 1/8 – je pense que tout commentaire sera superflu. Faute d'avoir traité cette question, les sociobiologistes ont chargé leur théorie d'une part considérable de mysticisme. » (SAHLINS 1980/1976, p. 92)

Un autre exemple de mécompréhension porte sur des affirmations de type « l'altruisme évolutionnaire est en réalité de l'égoïsme génétique ». Cette phrase choquante doit être décodée pour perdre son venin : « l'altruisme évolutionnaire » correspond aux comportements défavorables du point de vue de la *fitness* individuelle ; « l'égoïsme génétique » correspond à la définition vue ci-dessus ; le verbe « est » ne doit pas être compris au sens d'équivalence mais au sens d'explication. Richard DAWKINS par exemple se laisse parfois emporter dans le langage métaphorique si bien qu'il en vient à des thèses apparemment contradictoires. Tantôt il affirme « il apparaît souvent que l'acte apparemment altruiste n'est en réalité qu'un acte égoïste bien déguisé » (DAWKINS 1996/1976, p. 22), tantôt il propose de montrer « comment

En fin de compte, il est possible d'affirmer sans contradiction que l'altruisme évolutionnaire existe tout en admettant qu'il peut uniquement avoir été sélectionné s'il s'est avéré avantageux pour les gènes qui induisent ce comportement.

Pour savoir si un individu est altruiste par rapport à d'autres individus, on compare le nombre de descendants (plus précisément, ceux qui parviennent à l'âge adulte) des individus considérés au terme de leur cycle de vie ; si l'individu focal possède une moins bonne *fitness* que les autres *et* que cela est dû au fait qu'il a adopté un comportement d'aide, alors on peut dire de cet individu et de son comportement qu'ils sont altruistes. De cette manière, sont à considérer comme altruistes, les abeilles kamikazes, les marmottes siffleuses ou les vers du cerveau.

Au terme de ces pages consacrées au monde animal, nous n'avons passé en revue qu'une partie des théories qui permettent de rendre compte de l'altruisme évolutionnaire. Cela tient à ce qu'il existe un excellent candidat aux actions altruistes : l'homme. Or ce dernier doit être traité comme un cas à part, car il a acquis un grand nombre de compétences extraordinaires au fil de l'évolution, si bien que les théories proposées jusqu'à maintenant s'avèrent largement insuffisantes pour analyser son comportement. Les prochaines sections seront consacrées à l'altruisme propre aux êtres humains.

2.3. L'altruisme évolutionnaire propre aux êtres humains

2.3.1. La complexité de l'altruisme humain

Les êtres humains ont également leurs kamikazes (bien sûr d'abord les aviateurs japonais durant la Seconde Guerre mondiale) et leurs héros altruistes comme Winkelried ou Mère Teresa.[129] Plus généralement, nous sommes quotidiennement

---

l'égoïsme et l'altruisme individuels s'expliquent grâce à la loi fondamentale que j'appelle l'égoïsme des gènes » (DAWKINS 1996/1976, p. 24). Dans le premier cas, DAWKINS semble nier l'existence de l'altruisme, mais deux pages plus loin, on comprend qu'il cherche plutôt à l'expliquer par le recours à la perspective du gène.

[129] Ces exemples peuvent paraître fourvoyants dans la mesure où ils sont généralement utilisés dans le contexte des discussions sur l'altruisme psychologique (qui fera l'objet du chapitre 3). Pour ce qui nous

témoins ou auteurs d'innombrables formes d'investissements personnels en faveur du bien d'autrui. Il y a le soutien porté en faveur des membres de notre famille ;[130] il y a les grands et petits dons aux personnes démunies et associations caritatives ; il y a l'engagement personnel en faveur de causes en tout genre (en Floride par exemple, il est courant de voir des entreprises, institutions, voire des associations d'étudiants s'organiser volontairement pour entretenir des tronçons d'autoroute). Demandons-nous si les théories considérées jusqu'à maintenant permettent d'expliquer ces cas, au moins apparents, d'altruisme évolutionnaire.

A priori, si on admet que nous sommes un produit de l'évolution, tout porte à penser que notre capacité d'agir de manière altruiste est en partie le résultat de la pression de la sélection naturelle. Beaucoup de théoriciens évolutionnistes pensent que c'est au temps de l'émergence de l'*homo sapiens* (paléolithique supérieur)[131] que les comportements altruistes tels que nous les pratiquons aujourd'hui sont apparus et se sont stabilisés (BOEHM 1997 ; RICHERSON & BOYD 2000). Les études de paléontologie montrent qu'à cette époque, l'*homo sapiens* évoluait en petits groupes de chasseurs-cueilleurs dans des conditions de vie difficiles ; les membres de ces groupes se trouvaient donc dans une situation de dépendance mutuelle et de contacts répétés (STERELNY 2006). D'autre part, il est très probable que c'est à ce moment, que l'*homo sapiens* a développé des facultés mentales surprenantes par rapport aux autres animaux, lui permettant par exemple de reconnaître les individus déjà rencontrés et de se souvenir des interactions précédentes. Si ces hypothèses sont vraies, alors toutes les conditions nécessaires au fonctionnement de formes évoluées de sélection de parentèle[132], réciprocité[133] ou de signal coûteux étaient remplies. Dès lors, se pose la question de

---

occupe ici, il ne faut pas perdre de vue qu'il s'agit de savoir i) si des comportements observables de type « kamikaze » ou « mère Teresa » s'avèrent réellement désavantageux (du point de vue de la fitness) pour les individus qui les pratiquent tout en étant avantageux pour autrui, et ii) comment ces comportements ont pu évoluer.

[130] Voir par exemple Adam SMITH 2003 /1759, Section II, Introduction, pp. 305-307.

[131] Il est même possible que l'*homo sapiens* soit apparu dans un passé encore plus loin (à ce propos, voir MCBREATY & BROOKS 2000).

[132] L'altruisme envers les proches parents était certainement pratiqué bien avant le temps du paléolithique supérieur ; mais de nouvelles formes de sélection de parentèle qui nécessitent des capacités cognitives développées pour prodiguer une aide discriminative ont dû apparaître plus récemment.

[133] Au contraire des autres animaux, les êtres humains disposent de capacités cognitives suffisantes pour remplir toutes les conditions nécessaires à l'altruisme réciproque définies par TRIVERS (section 2.2.2.i).

savoir si ces théories, à elles seules et telles que nous les avons considérées jusqu'à maintenant, sont en mesure d'expliquer tous les cas d'altruisme (apparent ou non) chez les êtres humains.

Qu'en est-il de la sélection de parentèle ? Il est difficile de nier que nous sommes généralement plus enclins à aider nos proches parents ou les êtres qui nous ressemblent plutôt que les inconnus. Ce phénomène est certainement un effet du mécanisme de la sélection de parentèle, qui opère sur des comportements génétiquement déterminés. Certes, mais la sélection de parentèle fournit une explication bien trop grossière pour rendre compte du comportement humain. Les missions suicides des kamikazes japonais par exemple apparaissent dans un contexte culturel bien précis et on ne peut pas dire que ce phénomène est un simple effet phénotypique de notre code génétique. Il est donc illusoire de penser pouvoir réfléchir exclusivement en termes d'avantages génétiques pour expliquer l'apparition de ces comportements.

Ainsi, même si on peut accepter que nous n'échappons pas complètement à la sélection de parentèle, il faut admettre que les causes de nos comportements (népotiques ou autres) sont plus complexes. L'explication doit notamment tenir compte de notre capacité, très développée, de transmettre et assimiler des pratiques, des règles de comportements et des connaissances. De manière plus générale, elle doit également rendre compte de nos facultés, apparemment uniques, d'apprentissage et de raisonnement sur des choses complexes.

En bref, il semblerait que nos comportements résultent de trois facteurs. Ils sont influencées par une composante innée (par exemple les gènes qui induisent une tendance à vouloir venir en aide à ses proches parents), une composante d'apprentissage individuel (par le biais de leur expériences quotidiennes, les gens acquièrent et modifient leurs croyances, capacités, préférences, émotions, stratégies, en interagissant avec et en recevant un feedback de leur environnement) et une composante d'apprentissage culturel (en recevant un enseignement ou en imitant les autres, les gens forment en leur esprit des états mentaux similaires à ceux des autres).[134] Ainsi, pour une explication évolutionnaire d'un comportement propre aux êtres humains, il faut utiliser

---

On comprend désormais pourquoi les êtres humains sont de bien meilleurs coopérateurs que les animaux ! En revanche, on peut s'attendre à ce que beaucoup de comportements humains à première vue altruistes ne soient en réalité que des formes d'investissement égoïste à long terme.

[134] Cette catégorisation est due aux anthropologues évolutionnistes Robert BOYD et Peter RICHERSON (1985).

des modèles qui intègrent ces deux nouvelles composantes (l'apprentissage individuel et l'apprentissage culturel) et saisir la nature des interactions entre les forces génétiques, environnementales et culturelles. Or, ce n'est pas une mince affaire ! Nous verrons plus loin qu'il existe néanmoins des tentatives qui méritent d'être mentionnées.

Pour revenir à la sélection de parentèle, elle nous permet d'éclairer partiellement des phénomènes comme celui de l'aide prodiguée à nos proches parents ou notre propension à préférer les individus qui nous ressemblent. Pour ce qui est du deuxième cas, l'anthropologue Michael ALVARD (2003 ; voir aussi MCELREATH *et al.* 2003) explique par exemple comment les marques spécifiques d'appartenance à un groupe (coutumes vestimentaires, langue parlée, etc.) fournissent des indications sur la propension des individus à agir de manière coopérative ; ceux qui possèdent les mêmes marques s'identifient les uns aux autres et savent qu'ils partagent plus ou moins les mêmes normes et types de comportements. Cette connaissance va les pousser à se choisir mutuellement comme partenaires de coopération. Il s'agit ici d'un phénomène de sélection de parentèle élargie fortement imprégnée d'éléments culturels. Toutefois, la sélection de parentèle reste à première vue muette face à d'autres exemples de tendances au comportement altruiste évolutionnaire humain tels que la bravoure de Winkelried et des kamikazes japonais, l'engagement humanitaire de Mère Teresa ou les sacrifices que beaucoup de personnes sont disposées à faire au profit de grandes collectivités. Nous verrons plus loin (section 2.3.6) que certains psychologues évolutionnistes font précisément appel à la sélection de parentèle pour expliquer ce type d'altruisme. Mais avant cela explorons le potentiel des théories de la réciprocité et du signal coûteux.

Pour ce qui est des engagements en faveur de la collectivité et des petits sacrifices quotidiens au bénéfice d'autrui, on peut imaginer qu'il s'agit de l'application de stratégies coopératives dont on peut expliquer l'évolution au moyen des théories de la réciprocité ou du signal coûteux (tout en gardant à l'esprit qu'en fin de compte cela se résume à du « bon placement » à long terme qui rapporte un avantage au niveau individuel). Ces stratégies, lorsqu'elles sont utilisées par les êtres humains doivent être en partie génétiquement déterminées, en partie influencées par l'environnement culturel et en partie le résultat d'un calcul rationnel des bénéfices. Toutefois, les modèles itérés que nous avons considérés jusqu'à maintenant n'intègrent pas la possibilité de modifier sa stratégie au cours d'une partie (c'est-à-dire qu'ils modélisent uniquement des stratégies génétiquement déterminées). D'autre part, ces modèles ne présentent qu'une

image très imparfaite de la complexité des interactions humaines. Pour ce qui est de l'analyse de *Donnant Donnant* par exemple, on modélise uniquement des séries d'interactions binaires entre deux joueurs qui se rencontrent au hasard, conditions bien trop contraignantes pour représenter la réalité. En effet, au moyen de modèles mathématiques, BOYD et RICHERSON (1988) ont montré que l'altruisme réciproque fonctionne bien uniquement dans des contextes de relations dyadiques. En revanche, cette stratégie n'est plus efficace dans des contextes collectifs : dans un dilemme du prisonnier à n participants (où $n > 2$) par exemple, *Donnant Donnant* n'est pas stable. Plus généralement, l'altruisme réciproque s'avère inefficace dès qu'un groupe dépasse le nombre de personnes dont les individus sont capables de se souvenir ou lorsque, du fait de la grandeur du groupe, les individus n'interagissent pas très régulièrement. Or, les êtres humains interagissent souvent dans le cadre de grands groupes. Il faut donc se tourner vers des modèles plus raffinés, capables de simuler la richesse des interactions humaines.

2.3.2. La seconde génération de la théorie des jeux

Entrons dans l'ère de la seconde génération de la théorie des jeux avec des modèles développés en anthropologie évolutionniste et en économie expérimentale. Dans ce domaine, un bon nombre de chercheurs tentent de représenter des situations sociales complexes correspondant à la manière dont les êtres humains interagissent. Leur but est d'expliquer les conditions d'évolution et de stabilisation des comportements coopératifs et altruistes évolutionnaires humains. Leur programme se développe sur trois fronts : l'expérimentation et l'observation empirique, la modélisation sur ordinateur et l'explication théorique en termes évolutionnaires.

Dans le domaine empirique, des situations sociales sont simulées sous forme de jeux complexes, auxquels on fait participer des sujets humains. En guise d'illustration, considérons un jeu communément pratiqué : celui du « bien commun ». Il s'agit d'un type de dilemme social qui, depuis qu'il a été formulé en 1968 par Garrett HARDIN, a été largement étudié, notamment en économie et en anthropologie évolutionniste. Dans ce jeu, les sujets d'un groupe disposent chacun d'une même somme de départ et ont la possibilité de participer financièrement à un bien public ; chacun est libre d'investir le montant qu'il désire ou de ne rien donner du tout. L'ensemble des dons est ensuite

réuni, doublé puis redistribué à part égale à chacun des joueurs (indépendamment du fait qu'ils aient contribué ou non au bien commun). Ainsi, le rendement est le meilleur lorsque tous les participants investissent la totalité de leur somme de départ car les montants non investis ne sont pas doublés. Par contre, du point de vue de l'individu, l'investissement comporte le risque de recevoir en retour une somme inférieure à celle investie ; en effet, il est rationnellement plus avantageux de garder sa somme de départ tout en recevant sa part des différentes sommes investies par les autres joueurs dans le bien commun, et si la plupart des joueurs réfléchissent de cette manière, malheur à ceux qui investissent ![135]

Les expériences (MARWELL & AMES 1981; FISCHBACHER *et al.* 2001 ; voir aussi OSTROM 1998) pratiquées sur des sujets humains mis en situation de jeu de bien commun ont pu montrer que, contrairement aux prédictions de la théorie des jeux traditionnelle, beaucoup de sujets ont tendance à chercher la coopération, alors même que du point de vue de l'intérêt rationnel, cette stratégie est une erreur. Ainsi les gens sont plus généreux que ce que l'on pourrait penser ; il semblerait qu'ils soient altruistes.[136]

Un autre jeu que l'on trouve souvent dans la littérature est le jeu de la confiance (*trust game*). Le déroulement en est le suivant. Un premier joueur, le *truster* (celui qui fait confiance), reçoit 10 unités monétaires et peut décider quelle part de sa fortune il va donner à son partenaire. L'expérimentateur double ce montant (selon les jeux, la somme est triplée, voire quadruplée) puis le transfère au deuxième joueur, le *trustee* (celui à qui l'on fait confiance). Celui-ci décide s'il va donner quelque chose au *truster*, et si oui, combien. Son don éventuel est également doublé. Si les joueurs sont égoïstes, ils ne vont rien donner, car la tentation est grande pour le *trustee* de ne pas renvoyer l'ascenseur et de profiter ainsi d'un gain optimal. Pourtant, le résultat des expériences

[135] Au fond il s'agit d'un dilemme assez proche de celui du prisonnier avec la différence qu'il implique plus de deux participants.

[136] Toutefois, si l'on procède à un jeu du bien commun itéré où les individus jouent plusieurs coups de suite et cumulent leurs gains au fil des parties, on constate qu'au fil des jeux répétés, la coopération décline jusqu'à disparaître. Ce phénomène est probablement dû, d'une part, à l'existence d'opportunistes dont le comportement a pour effet de miner le désir de coopération des autres et d'autre part, à un biais assez cocasse chez les individus désireux de coopérer : ils attendent, de la part des autres, un taux de coopération légèrement supérieur au leur et se retrouvent donc régulièrement déçus dans leurs attentes ; puis en réaction, ils baissent leur taux d'investissement au coup suivant (FISCHBACHER *et al.* 2001).

montre le contraire : plus de 50% des *trustees* donnent en retour et leur contribution est proportionnelle à l'investissement du *truster*. Plus celui-ci est généreux, plus le *trustee* le récompense (FEHR & FISCHBACHER 2003 ; FEHR & ROCKENBACH 2003 ; HENRICH *et al.* 2004).

Ces résultats empiriques nécessitent une explication théorique. Il faut pouvoir rendre compte de cette tendance à coopérer qui est si forte chez les êtres humains. Si l'on veut recourir à une explication évolutionnaire, cette tendance est forcément liée à une stratégie dont il faut pouvoir montrer l'efficacité évolutionnaire. Dans le cadre du projet explicatif, les mêmes situations sociales sont alors modélisées sous forme informatique. Mais au contraire de ce qui se pratiquait dans la théorie des jeux traditionnelle (par exemple AXELROD), la conceptualisation informatique de ces jeux intègre tant bien que mal les paramètres de l'apprentissage individuel et culturel. Par exemple, au lieu d'avoir des individus qui répètent toujours la même stratégie (génétiquement déterminée) au cours d'une partie du jeu itéré, le modèle prévoit qu'une certaine proportion d'individus changeront de stratégie et copieront celle qui est la mieux représentée ; dans ce cas, la stratégie devient un élément culturel (HENRICH & BOYD 2001). Les nouveaux modèles sont également plus complexes. Ils ne modélisent pas uniquement des interactions dyadiques mais souvent des interactions triadiques ou *n*-adiques (BOYD & RICHERSON 1988); parfois, ils modélisent même des situations dans lesquelles les individus peuvent choisir (selon certains critères censés représenter la réalité) leurs partenaires de coopération (BRANDT *et al.* 2006). Enfin, les nouveaux modèles se caractérisent par le fait qu'ils font varier un grand nombre de paramètres. Par exemple, pour analyser les effets du contrôle social, tantôt la condition de l'anonymat est posée, tantôt elle est levée ; selon les situations, un individu pourra donc savoir, même s'il ne l'a jamais rencontré auparavant, si son partenaire d'interaction a agi précédemment de manière coopérative ou non.

Ces théories permettent d'expliquer une large palette de comportements apparemment altruistes en termes d'intérêt individuel sur le long terme. Elles montrent comment il peut être avantageux d'adopter une stratégie coopérative, même si elle prescrit des actions désavantageuses.

2.3.3. La réciprocité indirecte et le signal coûteux

La réciprocité directe est régulièrement pratiquée par les êtres humains dans des relations entre deux personnes. Elle permet d'expliquer un certain nombre de comportements d'aide. Mais comme nous l'avons vu, les conditions nécessaires à cette forme d'interaction sont très astreignantes (voir sections 2.2.2.i et 2.2.2.viii). Un autre moyen de garantir un certain taux de coopération et d'entraide tout en échappant à la condition de la répétition des rencontres entre deux individus est de pratiquer la réciprocité *indirecte*.

La théorie de la réciprocité indirecte se base sur l'idée que dans les sociétés humaines, les gens sont capables d'obtenir des informations au sujet d'autres individus (par le biais de l'observation et du commérage) et d'ajuster leur comportement en fonction des données qu'ils ont obtenu. En théorie des jeux, il est possible de modéliser ce genre de situations en levant la condition de l'anonymat. Selon la théorie de la réciprocité indirecte, il peut être intéressant de produire des actions apparemment altruistes en public afin de se forger une réputation qui incite les autres à entrer dans une relation de réciprocité (ALEXANDER 1987 ; PANCHANATHAN & BOYD 2004; FEHR 2004). Ainsi, au lieu de la devise « œil pour œil, dent pour dent » propre à l'altruisme réciproque, une stratégie de réciprocité indirecte suit la devise « aide ton prochain et tu seras aidé en retour ». Cela permet de rendre compte des situations où des êtres humains viennent en aide à des personnes dont ils savent pertinemment qu'ils ne pourront jamais rien attendre en retour.

La réciprocité indirecte est similaire au signal coûteux (section 2.2.3) ; un signal peut être un comportement public apparemment altruiste qui indique aux autres individus que l'on est un partenaire de coopération fiable. Il y a cependant une différence fondamentale entre les deux modèles. Le signal coûteux doit *avoir évolué* parce qu'il dévoile un trait non visible de l'individu porteur. Or la théorie de la réciprocité indirecte n'exige pas que le signal soit ancré dans nos gènes ; il peut simplement résulter d'une stratégie acquise par l'individu au cours de son ontogenèse et qu'il peut appliquer quand bon lui semble. La réciprocité indirecte est donc moins rigide que le signal coûteux.

La question reste ouverte de savoir si, dans les faits, les êtres humains donnent des signaux coûteux altruistes qui réfèrent à des dispositions correspondantes profondément ancrées en eux ; certains auteurs sont sceptiques (FEHR & FISCHBACHER 2003) et d'autres plus optimistes (GINTIS *et al*. 2001 ; G. ROBERTS 1998). Quant à la réciprocité indirecte, on ne peut nier qu'elle est particulièrement répandue dans notre espèce. Cela tient au fait que nous sommes extrêmement sensibles à la fois à notre réputation et à celle des autres. De manière générale, les expériences faites sur des êtres humains montrent que le comportement pro-social est fortement renforcé lorsque les joueurs disposent d'informations au sujets des comportements passés des autres joueurs ; nombre de nos décisions de prodiguer ou refuser l'aide à autrui dépendent de la réputation de l'autre (NOWAK & SIGMUND 1998 ; MILINSKI *et al.* ; 2001 ; 2002 ; WEDEKIND & MILINSKI 2000) de même que de l'état de notre propre réputation (SUGDEN 1986 ; LEIMAR & HAMMERSTEIN 2001 ; pour une revue de la littérature, voir MCELREATH *et al.* 2003).[137]

Ces modèles montrent à quel point la confiance est un élément important dans les interactions humaines (WEDEKIND & BRAITHWAITE 2002). Ils indiquent également que la punition semble être une condition essentielle pour le maintien de la coopération. Cette idée est renforcée par les expériences menées sur la punition altruiste à laquelle est consacrée la section suivante.

2.3.4. La punition altruiste

Les chercheurs de la seconde génération de la théorie des jeux ont montré que beaucoup de comportements humains paraissent altruistes évolutionnaires sans l'être en réalité. Des actions ponctuelles coûteuses pour l'individu peuvent être induites par une stratégie comportementale évolutionnairement non altruiste (qui s'avère avantageuse sur le long terme). Les théories de la réciprocité directe, indirecte et du signal coûteux permettent de comprendre le haut degré de coopération pratiqué par les êtres humains mais ce faisant, elles restreignent indéniablement le champ de l'altruisme évolutionnaire.

[137] Cette seconde condition marque le fait que si l'on veut se départir d'une mauvaise réputation, il peut valoir la peine d'aider un individu quelle que soit sa réputation (à ce propos, voir LEIMAR & HAMMERSTEIN 2001).

De manière assez intéressante, ce même groupe de chercheurs (GINTIS 2000 ; BOYD *et al.* 2003 ; FEHR & FISCHBACHER 2003) à mis en évidence un comportement altruiste évolutionnaire typiquement humain : la « punition altruiste ». Il s'agit d'un comportement justicier qui consiste à punir les opportunistes sans qu'il en résulte un bénéfice à long terme pour le punisseur.[138] Cette punition est altruiste, d'une part parce qu'elle engendre un coût (pour punir, il faut investir de l'énergie et des moyens), d'autre part parce qu'elle n'est pas liée à un retour de service ultérieur en faveur du punisseur.[139] De plus, ce comportement profite à d'autres individus car des tests empiriques ont montré que les opportunistes punis se comportent de manière nettement plus coopérative lors d'interactions ultérieures (FEHR & FISCHBACHER 2003). Enfin, la punition altruiste engendre des effets bénéfiques pour la coopération et l'entraide parce qu'elle force même les individus égoïstes à agir pour le bien d'autrui ; soudain, il vaut mieux être coopératif plutôt que de risquer la punition.

Des expériences ont montré que, dans les faits, les gens utilisent la punition altruiste dans des situations d'interaction sociale. Ernst FEHR et collègues (FEHR & GÄCHTER 2002 ; FEHR & FISCHBACHER 2004a) ont fait jouer des sujets humains à des variantes de différents jeux (dilemme du prisonnier, jeu de la confiance, jeu du bien commun, etc.) dans lesquels ils ont intégré le paramètre de la punition. Les résultats empiriques indiquent que s'ils en ont la possibilité, beaucoup de sujets sont prêts à punir les opportunistes à leurs propres frais tout en sachant qu'ils ne les rencontreront plus dans le cours du jeu (c'est-à-dire n'attendant aucun bénéfice en retour de leur punition).

---

[138] Dans un bon nombre d'écrits, on trouve ce comportement punitif associé au comportement d'aide à autrui pour former une seule stratégie : la « réciprocité forte » (*strong reciprocity*) (FEHR & GÄCHTER 1998; FEHR & FISCHBACHER 2003 ; GINTIS 2000). La réciprocité forte revient à afficher deux types de comportements en fonction du déroulement de l'interaction : récompenser la coopération et punir la non-coopération (dans les deux cas sans qu'il y ait retour de bénéfice ultérieur). Toutefois, il n'y a aucune raison de principe de considérer ces deux stratégies comme un tout indissociable (voir LEHMANN & KELLER 2006). Il se peut que ces traits aient évolué de manière indépendante (même s'il est clair que les deux favorisent la coopération). Pour cette raison, j'ai préféré ne pas parler de la réciprocité forte dans le corps du texte et traiter de manière séparée l'aide à autrui et la punition altruiste.

[139] C'est la raison pour laquelle des stratégies comme *Donnant Donnant* ne peuvent pas être considérées comme altruistes. Bien que coopérative et punitive, *Donnant Donnant* n'est pas altruiste car l'acte punitif n'engendre aucun coût supplémentaire (le punisseur refuse simplement de coopérer au coup suivant); au contraire, il permet ainsi d'éviter de se faire exploiter.

Sachant qu'au niveau individuel, il vaut mieux ne pas être altruiste puisque, par définition, les individus qui ne le sont pas s'en sortent mieux, comment peut-on expliquer que les comportements altruistes punitifs aient pu être sélectionnés au fil de l'évolution ? Il semblerait qu'il y ait différents facteurs complémentaires qui permettent d'expliquer la stabilisation des comportements altruistes punitifs.

Le premier facteur est celui du coût et de l'efficacité de la punition (BOYD & RICHERSON 1992). S'il y a suffisamment de punisseurs altruistes dans un groupe et que les punitions sont dissuasives (grand coût pour le puni), la coopération sera très répandue et les punisseurs altruistes devront rarement sévir, si bien que le coût engendré par leur comportement punitif sera moindre, voire nul. En comparaison des individus qui ne punissent pas, les punisseurs ne seront donc que légèrement désavantagés.[140] Toutefois, ce facteur à lui seul n'est pas suffisant pour assurer l'évolution des comportements altruistes punitifs. En effet, même si les coûts engendrés par le fait d'être un punisseur altruiste sont moindres, il n'en demeure pas moins qu'il vaut mieux être un simple coopérateur plutôt qu'un punisseur altruiste en plus ; le coopérateur non-punisseur profite des effets bénéfiques des comportements altruistes punitifs sans porter lui-même les coûts occasionnés lors de la punition des opportunistes. Dans ce contexte, on peut parler d'opportunisme de second ordre (*second order free riding*). Il est intéressant de noter ici qu'un effet de la punition altruiste est de transformer des traits hautement altruistes (ceux qui induisent des actions coopératives dans un monde d'égoïstes) en des traits à la fois avantageux du point de vue individuel (si la punition est efficace, il vaut mieux coopérer que faire défection)[141] et opportunistes de second ordre : dans un monde dominé par les altruistes punisseurs, la stratégie altruiste pure « Coopère toujours ! » peut être considérée comme opportuniste de second ordre.

L'opportunisme de second ordre nous force à chercher d'autres facteurs susceptibles de soutenir l'évolution de la punition altruiste. Notre second facteur est lié

[140] Evidemment, l'aspect « coût pour le punisseur » doit également être pris en compte ; pour que la pratique de la punition puisse se répandre, il faut que le coût pour le punisseur soit nettement inférieur au coût pour le puni.

[141] Lorsque la punition est extrêmement efficace, du point de vue des comportements, il n'y a plus moyen de distinguer entre les individus qui adoptent une stratégie coopérative uniquement dans un milieu punitif (afin d'éviter la punition) et ceux qui ont pour stratégie de toujours coopérer.

à la prescriptivité des comportements altruistes punitifs. Il repose sur l'idée qu'à un moment de l'histoire humaine, les normes sociales ont émergé.[142] Les normes sociales sont associées à une attente de comportements conformes à ce qu'elles prescrivent ; en cas de non-conformité, il y a sanction (OSTROM 1998). Pour renforcer les normes sociales, les êtres humains ont assigné une valeur prescriptive aux comportements punitifs eux-mêmes ; cette valorisation s'accompagne d'une *obligation* de punir, valable pour tous les membres du groupe.[143] Or si l'exécution de la punition devient un devoir, seront punis non seulement les opportunistes mais aussi les individus qui ne punissent pas (même si par ailleurs ce sont des coopérateurs) ; il y a donc punition des non-punisseurs, ou méta-punition.[144] Un bon nombre d'expériences théoriques et empiriques ont montré que le mécanisme de la punition des non-punisseurs renforce à la fois le comportement coopératif et le comportement punitif, c'est-à-dire qu'il permet de faire monter le taux moyen de coopération dans le groupe (BOYD & RICHERSON 1992 ; FEHR & FISCHBACHER 2004a).

Toutefois, la méta-punition a également ses limites : elle est confrontée à la difficulté d'une régression à l'infini car il vaut mieux être simple punisseur de non-punisseurs plutôt que punisseur de non-punisseurs de non-punisseurs, etc. Même s'il est clair qu'elle renforce la coopération et la punition, la méta-punition ne permet pas non plus à elle seule d'expliquer l'évolution des comportements altruistes punitifs.

Un autre facteur potentiel qui a fait couler beaucoup d'encre est celui de sélection de groupe (BOYD *et al*. 2003 ; GINTIS 2000 ; HENRICH & BOYD 2001). A la différence de la sélection génétique de groupe (section 2.2.4), ce qui va être présenté ici est une sélection *culturelle* de groupe où les objets de sélection ne sont pas des stratégies comportementales génétiquement déterminées mais des stratégies culturellement transmises. Voyons dans le détail comment elle fonctionne.

---

[142] La question de l'évolution des normes sociales et leur effet sur la coopération sera traitée à la section suivante (2.3.5).

[143] Dans ce contexte, certains auteurs parlent de *moralisation* des normes sociales et de la punition (GINTIS 2000). Cela me paraît toutefois exagéré. Que des règles soient valorisées et associées à la punition ne signifie pas que l'on entre dans le domaine moral. Ce point deviendra plus clair au chapitre 5, section 5.4.

[144] Cette idée de punition des non-punisseurs a déjà été élaborée en 1986 par Robert AXELROD.

On part du principe que les êtres humains forment des groupes relativement homogènes, composés d'individus qui adhèrent à des normes sociales et les transmettent par l'imitation et l'enseignement (section 1.2.1). En accord avec les observations ethnographiques, on admet également que les normes transmises diffèrent d'un groupe culturel à un autre : deux groupes voisins peuvent posséder des normes et des institutions très différentes. Ensuite, on présuppose que les êtres humains possèdent une « tendance au conformisme » (c'est-à-dire adoptent assez facilement les normes qui ont beaucoup de succès dans leur société) et une tendance à imiter les comportements qui ont du succès ou dont les individus qui les utilisent ont du succès (voir section 1.2.3, p. 43). Ainsi, au fil des générations, on observera à l'intérieur de chaque groupe, une tendance à l'uniformisation des normes acceptées (HENRICH & BOYD 2001 ; FEHR & FISCHBACHER 2003, p.790).

La sélection de groupe fonctionne s'il existe plusieurs groupes et si ces groupes sont suffisamment variés entre eux. Pour que ce soit le cas, deux conditions doivent être réunies. Premièrement, il faut une variation entre les normes sociales prônées dans les différents groupes ; par exemples les groupes peuvent se différencier par le fait que certains possèdent des normes sociales renforcées par la sanction et d'autres pas. Deuxièmement, il faut une influence de cette variation des normes sur la santé des groupes ; par exemple, on sait que les groupes qui possèdent des normes renforcées par la sanction se portent généralement mieux que ceux qui n'en ont pas, car ils sont plus efficaces dans la production de réserves, de moyens collectifs de défense, etc.

La sélection de groupe opère lorsqu'il y a compétition entre les groupes ; cette compétition se traduit par des guerres ou des conflits d'influence, qui se soldent soit par le dépérissement de certains groupes au profit des autres, soit par l'absorption d'un groupe par un autre ; dans ce dernier cas, les groupes vainqueurs imposent leurs normes culturelles et leurs institutions aux individus des groupes vaincus et le mécanisme du conformisme opère, au fil des générations, en faveur d'une uniformisation des normes sociales acceptées.[145] Ainsi, s'il y a compétition entre un groupe qui possède des normes renforcées par la sanction et un autre qui n'en possède pas, l'issue de la compétition se soldera par un avantage du premier sur le second. En conséquence, les comportements coopératifs et de punition altruiste se répandront dans l'ensemble de la population.

[145] On remarque ici la différence avec la sélection génétique ; un processus rapide d'uniformisation à l'intérieur des groupes contraste avec la rigidité de la transmission génétique.

En résumé, voici ce qui se passe pour le comportement altruiste punitif. Par définition, il est légèrement défavorable du point de vue individuel par rapport aux comportements non altruistes. Par contre, il se trouve qu'au niveau du groupe, l'existence d'individus altruistes est avantageuse puisqu'elle a pour effet d'augmenter la coopération qui permet la réalisation de projets communs d'envergure. Ainsi, si au niveau de la sélection individuelle le désavantage engendré par un comportement altruiste punitif n'est pas trop grand (les facteurs de l'efficacité de la punition et de la méta-punition agiront en ce sens), un petit effet de sélection culturelle de groupe suffit à faire pencher la balance à l'avantage de ce comportement (GINTIS 2000, p. 171).

Cette théorie de la sélection culturelle de groupe est un bel exemple de théorie de la coévolution gène-culture ; des tendances génétiquement déterminées influencent le processus d'évolution culturelle et il est probable que l'évolution culturelle de groupe a pour conséquence, sur le long terme, d'ancrer dans nos gènes des tendances psychologiques motivant à agir en faveur d'autrui (BOWLES *et al.* 2003).[146] De plus, la théorie de la sélection culturelle de groupe est plus crédible que son pendant génétique pour au moins deux raisons : d'une part, au niveau culturel, les groupes se font et se défont plus rapidement ; d'autre part, il est probable qu'à l'intérieur des groupes, la configuration des stratégies s'homogénéise assez rapidement (notamment grâce à la tendance au conformisme ou à la punition des non-punisseurs). Toutefois, elle implique une conséquence assez dérangeante : cette théorie fait dépendre l'évolution de l'altruisme évolutionnaire de l'existence de conflits permanents entre les groupes. Dit crûment : seule la guerre permet l'altruisme. Pour échapper à cet aspect indésirable, il vaut sans doute la peine de poursuivre la recherche de facteurs susceptibles d'expliquer la stabilisation des comportements altruistes punitifs. Le champ des nouvelles spéculations est ouvert.[147]

[146] Cette question sera développée au chapitre 4, section 3.4.

[147] Une solution serait de recourir à la réputation. Andy GARDNER et Stuart WEST (2004) par exemple soutiennent que l'ingrédient crucial pour l'évolution de la punition (et avec elle de la coopération) est une corrélation entre la stratégie punitive adoptée par un individu et la coopération qu'il reçoit en retour. Par exemple, si un individu est connu pour ses propensions à punir les opportunistes, par peur de la punition, les autres auront tendance à choisir de coopérer avec lui. Mais dans ce cas, la punition ne peut plus être qualifiée d'altruiste ! D'autre part, même en admettant l'impossibilité de l'évolution de la punition altruiste, cette explication n'est pas entièrement convaincante car elle dépend de contextes où la condition

Il va sans dire que de par leur aspect spéculatif, les modèles de la seconde génération de la théorie des jeux doivent être considérés avec précaution.[148] De plus, quoique déjà très raffinés, ils ne représentent que de manière imparfaite la complexité des relations humaines. Nous ne sommes par exemple jamais entièrement opportunistes et distribuons nos bienfaits et nos punitions de manière conditionnelle en fonction de critères plus complexes que la simple connaissance des actions passées des autres individus. Or ces finesses peuvent difficilement être modélisées. Ainsi, la théorie des jeux ne peut au mieux que révéler le fonctionnement et les implications de mécanismes très généraux.

Cela dit, on ne peut nier un apport important des études menées sur la réciprocité indirecte et l'altruisme fort : elles montrent à quel point la punition est un facteur crucial pour le maintien de la coopération.[149] Certains résultats sont sans appel : sans punition, la coopération cesse rapidement et lorsque la punition des opportunistes est possible, la coopération est plus stable (BOYD *et al.* 2003 ; FEHR & FISCHBACHER 2003).[150] De plus, la punition *altruiste* permet d'augmenter le taux de coopération, même dans le cadre de grands groupes (BOYD & RICHERSON 1992 ; FEHR & GÄCHTER 2002 ; FEHR & FISCHBACHER 2004a).[151]

---

d'anonymat est levée et le retour de service possible ; or les expériences montrent que beaucoup de sujets humains sont disposés à punir alors même que ces conditions ne sont pas remplies (GINTIS *et al.* 2003).

[148] Par exemple, il n'est pas évident de comprendre comment une stratégie punitive qui apparaît pour la première fois dans une population peut s'y propager ; en effet, même si la punition est peu coûteuse pour les individus qui la pratiquent dès lors qu'elle est suffisamment dissuasive, elle est en revanche très coûteuse pour les premiers altruistes punisseurs qui apparaissent dans une population peu coopérative. De nouveaux modèles ont récemment été développés pour résoudre cette question : FOWLER 2004 ; BOWLES & GINTIS 2004.

[149] Certains auteurs objectent que les comportements punitifs n'ont pas évolué parce qu'ils permettent de favoriser la coopération mais plutôt parce qu'ils ont pour effet de contraindre les opportunistes, c'est-à-dire de niveler les inégalités (M. PRICE *et al.* 2002). Il est difficile de savoir quel crédit accorder à cette objection, d'autant plus qu'il semble que la punition produise ces deux effets de manière conjointe. Peut-être faudrait-il considérer ces deux approches comme complémentaires en accordant une double fonction à la punition.

[150] Elle est cependant mise en péril lorsque les groupes deviennent trop grands (au-delà de 100 personnes).

[151] La question de savoir si la punition est un mécanisme qui permet l'évolution ciblée de comportements altruistes (FOWLER 2004) ou de n'importe quel comportement (BOYD & RICHERSON 1992) reste débattue.

2.3.5. Les normes sociales

Un élément culturel extrêmement intéressant lorsqu'on aborde la question de l'altruisme humain est celui des normes sociales. On trouve les normes sociales ainsi que les institutions qui les renforcent dans toutes les sociétés humaines et il paraît évident qu'elles soutiennent la coopération et la coordination ; elles jouent par exemple un rôle non négligeable dans l'explication de la punition altruiste.

Dès lors, il serait intéressant de disposer d'une explication de leur origine et de leur évolution. Voici une hypothèse qui méritera d'être complétée. Les premières normes sociales étaient probablement de simples conceptualisations de systèmes d'interaction préexistants. A un certain moment de leur évolution, les êtres humains ont acquis les capacités cognitives nécessaires à la compréhension de manière plus ou moins fine des effets bénéfiques des mécanismes de la réciprocité. Ils ont alors cherché à les appliquer de manière consciente. Mais cette prise de conscience par l'homme des bienfaits engendrés par la pratique commune des règles de réciprocité s'accompagne inévitablement d'un raffinement de l'aptitude à tricher.[152] Dans ces conditions, il faut une sorte de garde-fou qui permette de parer à l'égoïsme ponctuel et préserver les mécanismes de coopération (BOWLES *et al.* 2003 ; DEHNER 1998 ; voir aussi TRIVERS 1971[153] ; ALEXANDER 1987). Ce garde-fou est précisément l'élaboration de normes sociales liées à une clause d'obligation (et corrélativement à des sanctions). Le point important ici est de remarquer que les premières normes édictées devaient correspondre à des comportements sociaux évolutionnairement favorables à l'espèce.

---

[152] Un individu qui possède les capacités cognitives nécessaires pour comprendre les effets bénéfiques sur le long terme des pratiques coopératives, comprend également qu'il peut obtenir des gains directs ou s'épargner une dépense d'énergie en profitant de la coopération des autres. Or les avantages de la coopération disparaissent dès lors que trop d'individus décident d'agir contre les règles de coopération.

[153] TRIVERS à émis l'hypothèse que l'on trouve chez les êtres humains une évolution conjointe de formes de plus en plus sophistiquées de tricherie et de détection de la tricherie (voir chapitre 3, section 3.4.1). Cette « course aux armements » pourrait bien être une des fonctions biologiques principales du cerveau humain dans ses premiers balbutiements et un facteur important de son expansion. En d'autres termes, la course à la tricherie et à sa détection a eu pour conséquence, au fil de l'évolution, de développer et d'affiner les facultés mentales des êtres humains jusqu'à ce qu'ils aient acquis la capacité d'élaborer et agir en fonction de normes.

Les normes sont liées à des sanctions internes (émotions comme la culpabilité, la honte, le remords, la perte d'estime de soi, etc.) et externes (OSTROM 1998) ; ainsi l'acquisition des normes (ou plus précisément des capacités normatives) incite les individus à la fois à conformer leurs actions aux prescriptions des normes et à sanctionner les déviances à ces prescriptions.[154] Selon Herbert GINTIS (2003), la capacité de penser et agir de manière normative est l'expression d'une adaptation génétique (pour une analyse similaire, voir SRIPADA & STICH 2006).[155] Cette capacité aurait été sélectionnée parce qu'elle s'avère avantageuse du point de vue de la fitness individuelle ; en intégrant et se conformant à des normes, les individus peuvent mieux contrôler leurs pulsions, maintenir des relations interpersonnelles et élaborer des plans pour le futur. Quant aux normes produites par cette capacité, elles peuvent en principe être de nature égoïste ou altruiste (BOWLES *et al.* 2003), mais toutes ne pourront pas s'imposer dans une population. La sélection culturelle fera le tri parmi elles ; elle supprimera toutes celles qui sont trop désavantageuses pour les individus qui s'y conforment. La sélection culturelle *de groupe* peut par contre favoriser des normes neutres ou légèrement désavantageuses pour les individus qui s'y conforment : s'il y a compétition entre deux groupes dont le premier véhicule des normes sociales d'aide à autrui et le second non, il est hautement probable que le premier groupe l'emporte ; en effet, l'application de normes d'entraide renforce le groupe en facilitant la production de réserves communes, de moyens collectifs de défense, etc. (BOWLES & GINTIS 2004). Il n'y a donc rien de surprenant à ce que les normes qui prescrivent la coopération et l'entraide soient largement répandues dans les populations humaines. Il n'est pas étonnant non plus que ces mêmes normes soient souvent valables à l'intérieur du groupe mais deviennent caduques envers des individus étrangers (MOHR 1987 ; BOWLES & CHOI 2004).[156]

---

[154] C'est certainement en pensant à l'effet des normes de coopération que beaucoup d'auteurs associent les comportements altruistes punitifs aux comportements coopératifs, ce qui donne lieu à la stratégie de la « réciprocité forte » (voir note 138).

[155] « If an internal norm is fitness enhancing [position défendue par GINTIS], then the allele for internalization of norms is evolutionarily stable.» (GINTIS 2003, p. 418)

[156] Pour la petite histoire, Kenneth DOVER (1974) a étudié la moralité populaire au temps de la Grèce antique (par opposition aux théories morales élaborées par les philosophes de la même époque). Il apparaît que le comportement moral par excellence consiste à être bienveillant envers ses proches et amis *et* chercher à nuire à tous les autres ! (1974, p. 180)

En résumé, l'utilisation de normes sociales est apparue au cours de l'évolution parce qu'elle est liée à la prescription et à la sanction des déviances, ce qui permet de garantir la coopération et la coordination à l'intérieur de moyennes et grandes communautés, avec tous les avantages évolutionnaires que cela comporte (GINTIS 2003 ; BOWLES *et al.* 2003 ; GÄCHTER & FALK 2002 ; FEHR & FISCHBACHER 2003).

Cette explication est extrêmement sommaire et mériterait d'être étoffée davantage. Il se peut que les capacités normatives n'aient pas uniquement évolué pour restreindre la tricherie. Christopher BOEHM (2002/2000) par exemple pense qu'une cause majeure de leur évolution réside dans leur pouvoir de contrecarrer les abus des individus dominants, c'est-à-dire de garantir une certaine équité entre les différents membres de la société (à ce propos, voir LACHAPELLE *et al.* 2006).

De plus, le rapport entre le fait d'intégrer une norme et la motivation à agir conformément à cette norme (ou à punir les déviances) n'est pas suffisamment explicité. Dans le chapitre 5 (section 5.2.2) je défendrai l'idée que malgré les apparences, ce rapport, même s'il existe de manière indéniable, n'est pas de nature causale.

Enfin, au-delà de l'aspect punitif, il semblerait que l'utilisation de normes sociales favorise l'efficacité de la coopération lorsqu'elle est liée à un comportement d'aide discriminatoire envers les individus qui partagent les mêmes « marques ethniques » (*ethnic markers*) qui sont les traits visibles de leur appartenance ethnique (MCELREATH *et al.* 2003 ; ALVARD 2003). Les membres d'une ethnie partagent des traits culturels distinctifs : parmi ces traits, il y a leur langage, leurs habitudes culinaires, vestimentaires, etc. ; il y a également certains modèles de comportement et les normes sociales qui structurent les interactions. Les normes sociales permettent à deux individus d'une même culture de coordonner instantanément leurs interactions (généralement dans un sens coopératif). Grâce au ciment des normes sociales, aider de préférence les personnes qui présentent des marques ethniques similaires aux nôtres (dialecte, tenue vestimentaire) est une bonne garantie de retour de service à long terme.[157]

Au terme de ces réflexions, il apparaît que grâce à leur pouvoir de coercition et d'harmonisation des comportements, les normes sociales jouent un rôle clé dans

[157] Ce phénomène peut être considéré comme un effet barbe verte, c'est-à-dire un cas de sélection de parentèle élargie.

l'évolution de la coopération et de la coordination des interactions humaines. Il semblerait également que certaines d'entre elles induisent des comportements altruistes ; quoique la plupart des normes sociales sélectionnées au fil de l'évolution culturelle apportent des bénéfices aux individus qui les appliquent, dans certains cas, lorsque la sélection culturelle de groupe intervient, des normes sociales légèrement désavantageuses au niveau individuel peuvent se stabiliser à condition que leur application renforce le groupe.

2.3.6. Le retour de la sélection de parentèle

La punition altruiste et l'aide prodiguée en faveur des membres de notre communauté (en particulier les individus qui partagent les mêmes marqueurs ethniques) sont une chose, mais elles ne sauraient couvrir tous les types de comportements altruistes que l'on rencontre chez nos pairs.[158] Des personnes comme Mère Teresa ne sont ni des « punisseuses altruistes » ni de simples « bienfaitrices ethnocentriques » ; et les comportements d'abnégation au profit du groupe (à l'exemple de Winkelried) ne s'expliquent pas par l'application de normes sociales légèrement désavantageuses. Nous sommes donc toujours en mal d'une explication complète de l'altruisme évolutionnaire humain.

La solution a peut-être été trouvée par certains psychologues évolutionnistes. John TOOBY et Leda COSMIDES (1989) proposent une explication spéculative qui se base sur la théorie de la sélection de parentèle. Selon eux, cette dernière pourrait bien être à

[158] Notons également que dans le cadre de la théorie des jeux, les coûts et les gains sont toujours calculés en termes d'unités monétaires (ou d'autres formes équivalentes) : l'altruisme consiste à donner une partie de sa fortune ou renoncer à un gain, soit pour récompenser, soit pour punir. Il s'ensuit que lorsque ces modèles sont formulés de manière à représenter la dynamique de la sélection naturelle (en faisant se succéder plusieurs générations de parties itératives), on ne retrouve que la *fitness* classique (niveau de l'individu) dans sa composante « viabilité » ; la composante « fécondité » est calculée en fonction de la viabilité puisqu'on attribue une descendance proportionnelle aux gains obtenus par les individus au cours de la partie précédente. Ainsi la théorie des jeux ne peut pas modéliser le phénomène de sélection de parentèle qui repose sur la notion de *fitness* inclusive. Or nous avons vu que dans le monde animal, la sélection de parentèle est seule garante de l'existence de l'altruisme évolutionnaire. Nous verrons dans cette section que cette leçon n'a pas été oubliée par les psychologues évolutionnistes qui retournent à la sélection de parentèle pour expliquer l'évolution de l'altruisme humain.

l'origine des mécanismes proximaux[159] comme les émotions empathiques qui poussent les gens à agir de manière altruiste envers des individus non parents. Ces mécanismes proximaux auraient été façonnés au cours de la préhistoire humaine (probablement au temps du pléistocène) sous l'influence de la force de la sélection de parentèle, lorsque les êtres humains vivaient dans de petits groupes majoritairement constitués de proches parents. D'autre part, si, dans les conditions de vie en groupe, les êtres humains avaient peu de chance de rencontrer des individus non parents, il n'est pas nécessaire que les émotions empathiques soient dirigées de manière discriminatoire en faveur des proches parents au détriment des étrangers.[160] Au contraire, les interactions avec des étrangers étant rares, le coût nécessaire à l'acquisition du mécanisme de discrimination serait largement supérieur à celui engendré par des actions altruistes occasionnelles envers des individus non parents. C'est dans ce contexte précis que les mécanismes altruistes auraient été fixés dans notre matériel génétique. TOOBY et COSMIDES ajoutent que l'environnement dans lequel ces mécanismes ont évolué a subi des changements drastiques au cours des derniers millénaires ; aujourd'hui, nous vivons dans des groupes plus grands et extrêmement mobiles, si bien que ces mécanismes et le comportement altruiste qu'ils induisent ont probablement perdu leur vertu adaptative (voir aussi MOHR 1987).

Cette théorie est spéculative et repose sur un certain nombre de présupposés qui ne satisfont pas tout le monde (voir notamment SESARDIC 1995). Elle suppose par exemple qu'au pléistocène, les groupes étaient majoritairement constitués de proches parents ; mais c'est quelque chose qu'il est assez raisonnable de penser. De plus, cette théorie dissocie le processus de sélection de parentèle de la capacité de distinguer les proches parents puisqu'il est clair que les hommes du pléistocène possédaient déjà cette capacité ; ainsi, pour qu'elle soit crédible, il faudrait montrer que *le fait d'utiliser* la capacité de distinguer les individus pour décider de la manière dont on distribue nos bienfaits est évolutionnairement coûteux. Enfin, elle postule que les êtres humains vivaient en vase clos dans leurs petits groupes de membres apparentés. Or si c'était

[159] Les mécanismes proximaux sont, à l'échelle des individus, les causes directes des comportements. Pour une explication détaillée, voir p. 157.

[160] Ici, il faut distinguer entre le fait de posséder une capacité (en l'occurrence celle de reconnaître ses proches parents) et le mécanisme qui fait appel à cette capacité (on l'occurrence le mécanisme de discrimination en faveur des proches parents).

effectivement le cas, la force de sélection de parentèle devait être passablement atténuée par l'effet inverse de la compétition à l'intérieur du groupe (voir section 2.2.1.i, p. 60 et suiv.). On se débarrasse toutefois du problème si l'on peut montrer que nos ancêtres vivaient dans un environnement non saturé dans lequel ils pouvaient s'étendre (LEHMANN *et al.* 2006).

Conclusion

Au terme de ce chapitre, force est d'admettre que le problème de l'altruisme évolutionnaire est loin d'être résolu. D'importantes avancées théoriques ont déjà été faites mais d'autres se font encore attendre.

De manière générale, nous avons vu qu'il existe différents mécanismes sélectifs (sélection de parentèle, réciprocité directe, réciprocité indirecte, signal coûteux et éventuellement sélection de groupe) qui permettent l'évolution de la coopération et des comportements d'aide. Mais on ne peut pas simplement identifier ces derniers à l'altruisme. La plupart de ces comportements ne sont altruistes qu'au premier abord, car une fois déterminées les causes de leur évolution, on comprend qu'ils sont finalement avantageux pour les individus qui les pratiquent.

Chez les espèces animales, il semble que seul le mécanisme de la sélection de parentèle (compris au sens large de *fitness* inclusive) permet l'évolution de l'altruisme. Chez les êtres humains, en plus de la sélection de parentèle (dont l'effet est adapté dans certains cas et n'est plus, dans d'autres, qu'un vestige d'une ancienne adaptation), il est plausible que certaines formes d'altruisme (punition altruiste et comportements d'aide en faveur des membres de la communauté) soient apparues sous l'effet de la sélection culturelle de groupe qui opère sur les normes sociales. Mais cette théorie est encore jeune et controversée (LEHMANN & KELLER 2006).[161] Des recherches futures nous permettront sans doute de décider de sa pertinence.

[161] Dans un récent article, Laurent LEHMANN et Laurent KELLER (2006) résument la situation en proposant l'idée que les comportements d'aide peuvent évoluer si au moins une des quatre conditions suivantes est réalisée : i) un bénéfice direct pour l'agent, ii) la production d'une information sur le caractère coopératif de l'agent, ce qui favorise les relations de réciprocité directe ou indirecte (réciprocité directe ; réciprocité indirecte ; signal coûteux), iii) une haute probabilité d'interaction entre individus génétiquement proches (sélection de parentèle au sens strict), iv) une corrélation génétique entre les gènes responsables de l'altruisme et des effets phénotypiques identifiables (sélection de parentèle élargie ; effet

Enfin, un élément qui ressort de l'analyse de coopération humaine est l'importance de la punition. Si elle ne permet pas d'expliquer à elle seule l'évolution de comportements coopératifs (car ce n'est pas un mécanisme au même titre que la sélection de parentèle ou l'altruisme réciproque), elle lui apporte en revanche un sérieux coup de pouce en modifiant les rapports coûts-bénéfices au point où il devient avantageux de coopérer plutôt que d'être opportuniste.

« barbe verte »). Parmi ces quatre conditions, seules les deux dernières concernent l'évolution de l'altruisme évolutionnaire. On peut toutefois se demander s'il est vraiment utile de distinguer iii) de iv) puisque, comme nous l'avons vu (p. 63), l'effet barbe verte n'est qu'un cas particulier de la sélection de parentèle élargie (ce que LEHMANN et KELLER admettent d'ailleurs parfaitement). Notons également qu'ils refusent d'ajouter la sélection culturelle de groupe à leur liste de conditions suffisantes pour l'évolution de la coopération ; cette question reste débattue.

3. L'altruisme psychologique

Au chapitre précédent, nous avons parcouru une série d'explications relatives à l'évolution de l'altruisme évolutionnaire. Un comportement est altruiste s'il a pour *effet* d'augmenter la *fitness* (valeur de survie et de reproduction) d'un ou plusieurs autres individus aux dépens de la propre *fitness* de l'agent. Mais il faut noter que cette approche nous fournit peu d'indications sur l'altruisme tel qu'il est conçu par le sens commun tout comme en psychologie ou en philosophie. En effet, hors du monde de la biologie et de la théorie des jeux, les gens considèrent généralement qu'un acte est altruiste s'il résulte d'une *motivation* dirigée vers le bien d'un ou plusieurs autres individus. Or la notion de motivation, constitutive de ce que nous appellerons l'altruisme psychologique, est complètement étrangère à la définition de l'altruisme évolutionnaire.

A première vue, cette distinction nette entre altruisme évolutionnaire et altruisme psychologique est dérangeante car c'est plutôt la seconde forme d'altruisme qui est liée à la moralité. On peut donc se demander dans quelle mesure les considérations évolutionnaires du chapitre précédent peuvent être significatives dans le domaine moral. Ce chapitre a notamment pour objectif d'éclairer les liens entre ces deux formes d'altruisme et par ce biais, de montrer que l'altruisme évolutionnaire n'est pas complètement étranger au domaine moral.

Dans ce qui suit, je commencerai par proposer une définition de l'altruisme psychologique. Celle-ci permettra de saisir les différences et quelques points communs entre altruisme évolutionnaire et psychologique. Une fois ces distinctions établies, j'aborderai la fameuse controverse entre les défenseurs de la thèse de l'existence de l'altruisme psychologique et leurs opposants, partisans de la thèse de l'égoïsme psychologique; cette controverse pose la question de savoir si l'on peut réellement parler d'altruisme psychologique ou si ce qui passe pour tel n'est qu'une forme d'égoïsme déguisé. Pour commencer, je présenterai quelques arguments qui structurent ce débat mais nous verrons qu'ils mènent tous à une impasse : ils ne permettent pas de se décider pour un camp plutôt que pour un autre. Je proposerai de sortir de cette impasse en redéfinissant les termes de la controverse ; cette nouvelle perspective permettra à la fois de faire pencher la balance en faveur des défenseurs de l'existence de

l'altruisme psychologique, d'ouvrir le champ des spéculations sur les origines évolutionnaires de l'altruisme psychologique et de préciser les liens entre les formes psychologique et évolutionnaire de l'altruisme.

3.1. Une définition de l'altruisme psychologique

Voici une définition assez classique de l'altruisme psychologique largement inspirée de l'analyse du psychologue Daniel BATSON (1991, pp. 6-7).[162]

Une action est altruiste (psychologique) si elle est le résultat d'une motivation psychologique dirigée vers les intérêts et le bien-être d'autrui (et non vers les propres intérêts et bien-être de l'agent).

En d'autres termes, pour qu'une action altruiste psychologique puisse être réalisée, il faut une conjonction de trois phénomènes : il faut être conscient des intérêts et des conditions du bien-être d'autrui ; il faut être motivé à produire une action qui les réalise ; et enfin, il faut qu'aucune considération relative à nos propres intérêts et bien-être ne parvienne à contrebalancer cet élan en faveur d'autrui. Quelques précisions s'imposent par rapport à cette définition.

Tout d'abord, pour être rigoureux, il faudrait ajouter à cette définition une clause stipulant qu'une action altruiste doit être coûteuse pour l'agent. Saluer son voisin tous les matins parce qu'on l'apprécie et sait qu'il aime être salué peut difficilement compter comme action altruiste.

Pour ce qui est des termes de la définition proposée, la notion de *bien-être* doit être comprise de manière assez large pour inclure l'absence de souffrance physique, de sentiments négatifs (anxiété, stress, etc.) et la présence de plaisir physique et de sentiments positifs. Quant aux *intérêts d'autrui,* il ne faut pas les confondre avec ce que l'individu en question considère comme étant ses propres intérêts ; la définition prend

[162] « Altruism is a motivational state with the ultimate goal of increasing another's welfare. (…) Motivation is energy, a force within the individual (…) [which] is directed toward some goal (…) and [draws] the person toward this goal. (…) Dimensions of well-being include the absence of physical pain, negative affect, anxiety, stress, and so on, as well as the presence of physical pleasure, positive affect, security, and so on. » (BATSON 1991, p. 6)

en compte ce que l'agent altruiste considère comme étant les intérêts de l'individu qu'il aide. Enfin la *motivation psychologique* réfère à une force qui pousse à l'action ; cette motivation semble être liée à des émotions dirigées vers autrui comme l'amour, la compassion, la sympathie ou la pitié.

D'autre part, agir de manière altruiste signifie que l'on ne prend pas en considération nos propres intérêts et notre propre bien-être ou au moins que l'on y accorde une importance moindre par rapport aux intérêts et bien-être d'autrui. Par contre, cette mise à l'écart de notre bien-être et intérêts propres ne doit pas forcément se faire de manière consciente. Il me semble qu'une mère qui se jette spontanément dans une rivière pour sauver son enfant qui y est tombé par mégarde agit de manière altruiste psychologique même si avant d'agir, elle n'a pas mis consciemment de côté ses propres intérêts.[163] Il est même probable que cette mère ait agit de manière tellement intuitive et spontanée que l'on ne puisse même pas parler d'intention de sa part de sauver son enfant.[164] C'est pour cette raison qu'au contraire de certains auteurs (SESARDIC 1995, p. 129 ; KITCHER 1987/1985, p. 397), je ne voudrais pas définir l'altruisme psychologique en termes d'*intentions*.

Il est important de préciser ici que par définition, la motivation altruiste ne peut pas être de nature instrumentale. Il est possible d'être motivé à agir en faveur d'autrui en pensant qu'il s'agit d'un bon moyen pour réaliser en fin de compte notre bien-être personnel, mais dans ce cas, on ne peut pas parler de motivation altruiste.

Cela dit, une action altruiste n'est pas forcément un sacrifice de soi. Il peut arriver qu'au moment du choix de l'action, nous ne soyons pas conscients du fait qu'elle

[163] Notons que tous les philosophes ne seraient probablement pas disposés à utiliser le qualificatif d'altruisme psychologique pour des actions en faveur d'un ami ou d'un parent. Pour ce genre de cas, Charlie BROAD (1971/1953) par exemple parle d'altruisme autoréférentiel (*self-referential altruism*), préférant réserver l'altruisme authentique aux actions motivées pas des préoccupations universalistes pour autrui. Il me semble cependant que ce genre de distinctions complique inutilement le débat. De plus, si l'on veut proposer une explication évolutionnaire de l'altruisme psychologique (comme ce sera le cas plus loin dans ce chapitre), il faut éviter d'en donner une définition trop complexe. La définition doit être compatible avec une approche en termes de constantes psychologiques qui puissent faire objet de sélection. L'altruisme psychologique conçu simplement en termes de motivation psychologique dirigée vers le bien d'autrui fait bien l'affaire car il permet de focaliser l'attention sur des émotions comme l'amour ou la sympathie.

[164] Cela nous mène à la discussion de la possibilité d'actions qui découlent directement de réactions émotionnelles ; ce que l'on appelle en anglais les *actions out of emotions* (DÖRING 2003).

produira un bénéfice autant favorable à nous-mêmes qu'à autrui. Il peut aussi arriver que l'on évalue incorrectement une situation si bien que notre action, quoique découlant d'une motivation altruiste, s'avère en fin de compte avantageuse pour nous et désavantageuse pour autrui. Imaginons par exemple que Roger, en traversant le désert avec une réserve d'eau qu'il craint insuffisante pour ses propres besoins, offre généreusement une bouteille d'eau à un homme assoiffé qu'il croise au cours de son périple, sans savoir que le liquide est mélangé à un poison mortel qui lui était destiné…

Au vu de la définition présentée, il devient clair que produire des actions altruistes psychologiques exige certaines capacités cognitives. Il faut en tout cas avoir conscience d'autrui en tant qu'être différent de nous-mêmes et qui peut avoir des intérêts et des états psychologiques propres, des buts similaires à nos propres buts ; en d'autres termes, il faut posséder une conscience de soi et la capacité de la théorie de l'esprit.

La conscience de soi[165] réfère à la possibilité qu'a un individu de se constituer une représentation de ses caractéristiques plus ou moins permanentes et de porter son attention sur sa propre personne, c'est-à-dire de se prendre lui-même comme objet de pensée (à ce propos, voir CLEMENT 2007, p. 178). Des expériences ont montré que les enfants à partir de 18 mois (LEWIS & BROOKS-GUNN 1981) ainsi que les grands singes comme les orangs-outans ou les chimpanzés (GALLUP 1977) sont capables de s'identifier et se reconnaître.[166] Ainsi, même les animaux possèdent une forme minimale de conscience de soi.

La théorie de l'esprit est une intentionalité de deuxième ordre, une capacité de raisonner au sujet des états mentaux d'autres individus (PERNER & WIMMER 1985).[167]

[165] La *conscience de soi* doit être distinguée de la *conscience phénoménale* (état conscient qualitatif) et de la *conscience accès* (qui réfère au contenu informationnel de nos états mentaux conscients). A ce propos, voir Ned BLOCK 1995.

[166] Pour tester cette faculté, on fait souvent passer aux sujets le test du miroir qui consiste à appliquer à leur insu une marque rouge sur leur front, puis les placer devant un miroir ; si le sujet réagit à la vue de la tache sur son front, alors on peut admettre qu'il possède cette forme de conscience de soi.

[167] Il existe des explications concurrentes de l'évolution de la théorie de l'esprit. Michael TOMASELLO et collègues (2005) pensent que cette faculté a évolué dans le cadre de sociétés coopératives où les individus entreprennent des projets communs. De même, au niveau de l'ontogenèse, la théorie de l'esprit se développe lorsque les enfants commencent à entrer dans des dynamiques de coopération (H. MOLL & TOMASSELLO 2007). Contrairement à l'idée de la théorie de l'esprit comme adaptation pour la vie sociale coopérative, Joseph HENRICH et Richard MCELREATH (2003) soutiennent qu'elle a évolué parce qu'elle

Les êtres humains (à l'exception des autistes) sont généralement capables de ce genre d'états mentaux à partir de l'âge de quatre ans et les maîtrisent pleinement dès l'âge de six ans. Les spécialistes divergent sur la question de savoir si d'autres espèces animales possèdent la théorie de l'esprit. Peut-être que certains grands singes tel que les chimpanzés possèdent une capacité de lire dans l'esprit d'autrui ;[168] mais même à supposer que ce soit le cas, elle est nettement plus limitée que chez les êtres humains dès l'âge de six ans (pour une revue de cette abondante littérature, voir DUNBAR 2000 ; PROUST 2003). Ainsi, seuls les êtres humains et peut-être les grands singes sont capables de produire des actions altruistes psychologiques.

3.2. Altruisme évolutionnaire *versus* altruisme psychologique

Au vu de la définition proposée à la section précédente, il apparaît clairement que l'altruisme psychologique et l'altruisme évolutionnaire sont deux notions logiquement indépendantes ;[169] tout comportement altruiste évolutionnaire n'est pas forcément altruiste psychologique et inversement. Par exemple le comportement des abeilles kamikazes est clairement altruiste évolutionnaire sans pour autant être motivé par la prise en considération du bien-être de leurs consœurs. Il s'agit d'un comportement entièrement génétiquement déterminé qui ne nécessite aucun type particulier de motivation psychologique. Inversement, l'action altruiste psychologique de Roger qui offre généreusement sa bouteille à une personne assoiffée sans savoir qu'elle contient un poison mortel aura des effets désastreux sur la *fitness* de la personne assoiffée. Il s'agit d'un cas particulier d'altruisme psychologique sans contrepartie dans l'altruisme évolutionnaire.

---

apportait un avantage sélectif direct aux individus capables de faire des prédictions sur le comportement d'autrui. Prédire le comportement d'autrui permet d'acquérir et d'utiliser des informations pratiques dans le domaine de l'interaction sociale et d'ajuster son comportement de manière à favoriser ses intérêts propres.

[168] Contre des auteurs comme Daniel POVINELLI et collègues (1992), Henrike MOLL et Michael TOMASELLO (2007) contestent l'idée que les grands singes soient réellement capables de changer de rôle et prendre la perspective d'autrui (voir aussi TOMASELLO *et al.* 2003).

[169] Cette distinction logique a été très clairement mise en évidence par D. WILSON et SOBER (2003/1998 ; 2002/2000, p. 186). Ces deux auteurs s'insurgent contre l'amalgame fréquent entre altruisme évolutionnaire et altruisme psychologique.

Ce dernier exemple en particulier met en évidence une limite importante de la comparaison entre ces deux formes d'altruisme. Je m'explique. Un théoricien évolutionniste s'intéresse aux objets de sélection. Son attention se porte sur les raisons de l'apparition de ces objets au cours de l'évolution, sur leurs effets à long terme et sur les conditions de leur sélection ou stabilisation évolutionnaire. Ainsi tout objet de sélection se doit d'être réplicable et représenté de manière multiple dans une population. Un type de comportement régulièrement pratiqué ou une tendance psychologique peut par exemple faire office d'objet de sélection. Par contre ce n'est certainement pas le cas des actions particulières.

En philosophie, en revanche, on travaille souvent sur des exemples théoriques particuliers ; beaucoup de philosophes apprécient d'ailleurs les expériences de pensées irréalistes. Cette approche proprement philosophique permet d'analyser les actions particulières ; par exemple décider si l'action de Roger est le résultat d'une motivation altruiste ou non.

Si l'on veut utiliser ce type d'approche pour aborder le phénomène de l'altruisme évolutionnaire, on se trouve rapidement confronté à de grandes difficultés. Considérons un exemple. Imaginons que Max déteste Julie et désire sa mort. Disons que Julie apprécie les promenades dans la nature et emprunte régulièrement le même itinéraire. Sachant cela, Max élabore le projet suivant : il va se cacher aux abords d'un petit pont qui surplombe une rivière au courant rapide et tumultueux, attendre que Julie s'y engage et surgir soudainement en face d'elle pour la précipiter dans le vide. Ce que ni Julie ni Max ne savent, c'est que le pont est vermoulu et menace de s'écrouler sous le poids d'une charge humaine. Au cours de sa promenade, Julie s'engage lentement et tranquillement sur le pont. Ce dernier commence à céder sans qu'elle s'en rende compte. Au même moment, Max surgit de sa cachette et s'élance, depuis l'autre côté du pont, à l'encontre de Julie. Surprise, Julie fait un bond en arrière et se retrouve sur la terre ferme. Le pont s'écroule sous le poids de Max, le précipitant dans les flots. Si Max ne s'était pas élancé sur le pont en effrayant Julie, c'est elle qui se serait retrouvée au fond de la rivière. En adoptant une approche philosophique qui se concentre sur l'analyse de l'action particulière, l'action de Max apparaît clairement comme non altruiste du point de vue de la motivation. Or à l'opposé, la définition de l'altruisme évolutionnaire nous force à considérer cette action comme altruiste, ce qui paraît choquant.

Evidement, présentée de cette manière, la notion d'altruisme évolutionnaire ne peut être que discréditée ; mais cela tient uniquement au fait qu'elle est décontextualisée par une approche philosophique qui focalise sur les actions particulières. Il est essentiel de garder à l'esprit que l'altruisme évolutionnaire est pertinent dans un contexte évolutionnaire, où l'on s'intéresse à des types de comportements régulièrement pratiqués dans une population.[170]

On peut se demander pourquoi les biologistes et les philosophes font usage de la même notion. La raison tient à ce que le terme « altruisme » se construit sur une tension entre les intérêts d'autrui et les intérêts personnels et cette tension se solde généralement par un avantage pour les intérêts d'autrui au profit des intérêts personnels. Dit autrement, à la fois le sens évolutionnaire et le sens psychologique réfèrent au fait de promouvoir les intérêts d'autrui au détriment de ses propres intérêts. La différence réside en ce que les philosophes focalisent leur attention sur les motivations à l'action alors que les biologistes, anthropologues, économistes et théoriciens des jeux évolutionnaires considèrent les effets des actions (conçues au sens large). Le choix de ces derniers relève du fait qu'ils travaillent dans le domaine des sciences empiriques où l'on constate des faits et où l'on s'efforce de les expliquer ; au fond c'est leur méthode de travail qui les a poussés à s'éloigner de la notion d'altruisme telle qu'elle est utilisée dans le sens commun.

Le fait que ces deux définitions de l'altruisme partagent un certain point commun tout en étant différentes, mène souvent à des confusions dans l'esprit du lecteur lorsque les deux notions apparaissent dans un même écrit. Par exemple, dans « The Nature of Human Altruism » (2003), Ernst FEHR et Urs FISCHBACHER commencent par préciser que dans le cadre de leur écrit, ils utiliseront le terme « altruisme » dans le sens d'actions coûteuses qui confèrent des avantages économiques à autrui.[171] Puis à la même page, ils font soudain usage de la notion d'altruisme psychologique en parlant de motifs altruistes qui induisent des comportements coopératifs et punitifs altruistes (alors

[170] A la rigueur, elle pourrait être pertinente pour des philosophes qui s'intéressent aux dispositions ou tendances générales à l'action.

[171] « Throughout the paper we rely on a behavioural – in contrast to a psychological – definition of altruism as being costly acts that confer economic benefits on other individuals ». Cette définition est directement calquée sur celle de l'altruisme évolutionnaire, que les auteurs rendent ainsi : « (…) fitness-reducing acts that confer fitness benefits on other individuals » (FEHR & FISCHBACHER 2003, p. 785).

qu'aucun argument n'est présenté dans leur article pour soutenir cette affirmation).[172] Parfois les deux notions sont même intégrées dans la même phrase. Ainsi, dans un contexte où il est clairement question de comportements altruistes évolutionnaires, FEHR et Bettina ROCKENBACH en viennent à écrire que « les coopérateurs altruistes veulent coopérer (…) même si la défection leur serait plus avantageuse. » (2003, p. 137).[173] Ces écrits ne sont pas contradictoires à proprement parler mais ils intègrent, sans les distinguer suffisamment, deux niveaux de discussion. Pour cette raison, ils doivent être lus avec une extrême précaution afin de ne pas se tromper sur la portée réelle des arguments et explications présentés.

A ce stade de la réflexion, les lecteurs pessimistes pourraient penser que puisque les philosophes ne sont pas autorisés à utiliser leurs méthodes pour analyser la pertinence de la notion d'altruisme évolutionnaire, les théoriciens évolutionnistes feraient bien de ne pas se mêler d'altruisme psychologique ; cela éviterait bien des confusions et conflits inutiles. Je n'abandonnerais pas si vite le dialogue interdisciplinaire. A défaut de rapport logique, on peut au moins parler d'un lien de type fréquentiel. En effet beaucoup d'actions causées par des motifs dirigés vers le bien d'autrui ont pour effet d'augmenter le bien-être d'autrui. Ce lien fréquentiel ouvre la possibilité d'une coévolution des comportements altruistes évolutionnaires et d'une propension à avoir des motivations altruistes. Nous verrons plus loin qu'à y regarder de plus près, il semblerait que l'altruisme évolutionnaire soit une condition nécessaire à l'évolution de l'altruisme psychologique.[174] C'est du moins dans cette direction que se dirigent les travaux des biologistes, anthropologistes évolutionnistes et théoriciens des jeux. Mais avant de traiter cette question dans le détail, il faut commencer par répondre à une objection radicale selon laquelle l'altruisme psychologique n'existe tout simplement pas. Ce sera l'objet des sections suivantes.

---

[172] « (…) a combination of altruistic and selfish concerns motivates them [les réciprocateurs forts]. Their altruistic motives induce them to cooperate and punish in one-shot interactions and their selfish motives induce them to increase rewards and punishment in repeated interactions or when reputation-building is possible. » (FEHR & FISCHBACHER 2003, p. 785)

[173] « Altruistic cooperators are willing to cooperate, that is, to abide by the implicit agreement, although cheating would be economically beneficial for them » (FEHR & ROCKENBACH 2003, p. 137)

[174] Une amorce de solution a déjà été annoncée à la section 2.3.6 : TOOBY et COSMIDES (1989), en introduisant des mécanismes proximaux comme les émotions empathiques (qui causeraient l'altruisme évolutionnaire), mettent en jeu quelque chose de très ressemblant à l'altruisme psychologique.

3.3. La controverse entre altruisme et égoïsme psychologiques

3.3.1. Les termes de la controverse

Nous avons vu que selon la définition de l'altruisme psychologique, si une action est le résultat d'une motivation psychologique dirigée vers les intérêts et le bien-être d'autrui (et non vers les intérêts et bien-être de l'agent lui-même), alors elle peut être considérée comme altruiste.

Un certain nombre de penseurs défendent le point de vue que les êtres humains sont incapables de réaliser de telles actions car ils peuvent uniquement être motivés par des considérations relatives à leurs propres bien-être et intérêts. Je les appellerai les *partisans de l'égoïsme psychologique*. Parmi ceux-ci, on trouve des philosophes (HOBBES 2000/1651 ; MANDEVILLE 1990/1714) et des psychologues (CIALDINI *et al.* 1987 ; CABANAC *et al.* 2002). La variante la plus répandue de la thèse de l'égoïsme psychologique est celle de l'hédonisme psychologique, selon laquelle toutes nos actions sont motivées par des considérations relatives à nos propres plaisirs et peines (où l'on recherche le plaisir et fuit les expériences désagréables).

D'autres penseurs défendent la thèse inverse. Parmi eux, il y a notamment des philosophes (BUTLER 1991/1726 ; A. SMITH 2003/1759, p. 47; NAGEL 1970), des psychologues (BATSON 1991) et des biologistes ou philosophes des sciences (E. WILSON 1979/1978[175] ; SOBER & D. WILSON 2003/1998 ; 2002/2000). Adam SMITH par exemple écrivait :

[175] En réalité, la position d'Edward WILSON est un peu ambivalente. Il écrit bien « L'impulsion altruiste peut être irrationnelle et dirigée unilatéralement vers autrui ; l'altruiste n'exprime aucun désir de réciprocité et nulle composante inconsciente ne vient entacher la pureté de son acte. J'ai appelé cette forme de comportement l'altruisme 'pur' » (1979/1978, p. 227). Mais cette prise de position très claire en faveur de la thèse de l'altruisme psychologique est précédée de formulations qui semblent indiquer le contraire : « Aucune forme d'altruisme humain n'est explicitement et totalement suicidaire. Les héros les plus grands risquent leur vie dans l'espoir de grandes récompenses, dont la moindre n'est pas l'immortalité en laquelle ils croient » (1979/1978, p. 225); « La pitié est sélective, et en dernier ressort on en tire souvent avantage. » (1979/1978, p. 226)

> « Aussi égoïste que l'homme puisse être supposé, il y a évidemment certains principes dans sa nature qui le conduisent à s'intéresser à la fortune des autres et qui lui rendent nécessaire leur bonheur, quoiqu'il n'en retire rien d'autre que le plaisir de les voir heureux. » (A. SMITH 2003/1759, p. 23)

Ainsi, la controverse entre l'égoïsme et l'altruisme psychologique porte sur la possibilité de l'existence d'actions altruistes psychologiques.

Précisons d'emblée que la thèse de l'égoïsme psychologique est *descriptive* (elle ne dit rien sur ce qui devrait être fait), qu'elle n'implique pas forcément la malveillance et qu'elle n'exclut nullement que nos actions égoïstes (au sens psychologique du terme) puissent avoir pour effet de favoriser le bien-être ou les intérêts d'autrui. D'autre part, la thèse de l'altruisme psychologique ne nie pas la possibilité d'actions motivées par des considérations égoïstes. Elle se contente d'affirmer qu'il est possible d'être poussé à l'action par des motivations altruistes ou que s'il y a conflit entre des motivations égoïstes et altruistes, il arrive que les secondes l'emportent et soient la cause de l'action. La thèse de l'altruisme psychologique défend donc l'existence d'un pluralisme motivationnel (à ce propos, voir SOBER & D. WILSON 2003/1998, chap. 7). Elle est même compatible avec l'idée que l'égoïsme est largement répandu dans ce monde.

Dans ce qui suit, je vais présenter divers arguments provenant du camp des partisans de l'égoïsme psychologique ainsi que de leurs opposants. Aucune de ces tentatives ne me paraissant concluante, je proposerai de recadrer le débat en distinguant deux manières de concevoir l'altruisme psychologique : la version purement motivationnelle et la version sophistiquée qui réfère aux désirs, buts et intentions des gens. Je montrerai ensuite que si l'on s'attache à la version motivationnelle, les défenseurs de l'altruisme psychologique l'emportent dans la controverse. Dès lors que l'on donne du crédit à l'existence de motivations altruistes, surgit la question de l'explication de leur évolution. Je défendrai l'idée qu'une telle explication (du moins certaines formes de cette explication) peut être vue comme un argument en faveur de la thèse de l'altruisme psychologique.

3.3.2. Quelques arguments pour et contre

Dans le cadre de la controverse entre l'égoïsme et l'altruisme psychologique la stratégie utilisée par les partisans de l'égoïsme psychologique est habituellement extrêmement simple : elle consiste à trouver une explication en termes de motivation égoïste pour chaque situation ou type de situation apparemment altruiste. Quant aux défenseurs de l'altruisme psychologique, leur argumentaire est généralement plus varié. Cette section est dédiée aux arguments susceptibles de décider lequel des deux camps peut l'emporter : elle se conclura cependant sur un match nul.

Un argument provenant du camp des défenseurs de l'altruisme est de dire que la thèse de l'égoïsme psychologique présente une image peu reluisante de la manière dont les êtres humains réfléchissent et orientent leurs actions (JOYCE 2006, p. 48 ; JAMIESON 2002). Par exemple, selon la version hédoniste de cette thèse, l'ensemble de nos choix relève d'une seule dimension de notre pensée : les considérations sur notre propre plaisir.[176] Cette approche semble donc relever d'une conception bien cynique du comportement humain. Cet argument est toutefois assez faible puisqu'un partisan de l'égoïsme psychologique pourrait simplement rétorquer que la réalité ne correspond pas toujours à l'image que l'on s'en fait.

Une autre ligne d'attaque consiste à recourir à des données expérimentales pour faire pencher la balance d'un côté plutôt que de l'autre. A la section 2.3 nous avons vu que les êtres humains ne choisissent pas systématiquement les actions qui leur apportent un bénéfice. En faisant jouer des sujets humains à des variantes de différents jeux (dilemme du prisonnier, jeu du dictateur, jeu du bien commun) on constate que beaucoup de sujets coopèrent sachant parfaitement que l'action la plus avantageuse pour eux serait la défection (section 2.3.2); on constate aussi que les gens sont prêts à punir les opportunistes, à leurs propres frais et sans attente de bénéfices en retour (punition altruiste : section 2.3.4). Souvenons-nous également du jeu de la confiance où les sujets récompensent largement les actions coopératives (p. 117). De plus, une version légèrement modifiée de ce jeu a révélé des résultats encore plus étonnants. Dans cette

[176] « The fundamental problem with HE is that it does not adequately explain the actual choices that people make. It crudely conceptualizes people as simple, one-dimensional, decision-makers, seeking to realize only one value (pleasure). » (JAMIESON 2002, p. 707)

version, on ajoute un troisième joueur qui n'est en fait qu'un observateur auquel on donne une certaine somme de départ et qui peut, durant le jeu, punir les autres joueurs. Mais la punition comporte un coût et il est clair pour l'observateur qu'il n'obtiendra aucun gain quel que soit l'issue des interactions entre les autres joueurs. En bref, la seule chose que peut faire l'observateur est de dépenser son argent pour punir ou garder son argent en s'abstenant d'intervenir dans le jeu. En principe, si l'observateur était un individu égoïste et rationnel, il ne devrait punir personne pour garder jalousement son bien. Pourtant au cours des expériences qui ont été faites, il est apparu que deux tiers des observateurs punissent régulièrement les opportunistes (FEHR & FISCHBACHER 2004b).

Richard JOYCE (2006, p. 48) pense que des données empiriques de ce type peuvent être utilisées en faveur de la thèse de l'altruisme psychologique. Pour tester cette hypothèse, demandons-nous quels pourraient être les ressorts psychologiques sous-jacents aux tendances à récompenser les actions coopératives et à punir les actions opportunistes. S'il s'agit de motivations altruistes, alors les sujets devraient être motivés à faire du bien aux autres joueurs. Ils pourraient se sentir bienveillants envers les autres joueurs, ce qui les motiverait à agir de manière coopérative et à punir les opportunistes afin qu'ils cessent de profiter de ceux qui coopèrent (la punition serait alors un bienfait indirect). Toutefois, cette interprétation peut être attaquée sur plusieurs fronts.

Un défenseur de l'égoïsme psychologique pourrait rétorquer que malgré les apparences, la tendance à coopérer et à récompenser les actes de confiance s'avère en fin de compte égoïste du point de vue de la motivation. Pour soutenir cette thèse, il pourrait évoquer les travaux de James RILLING et collègues (2002). Dans des études sur les bases neuronales sous-jacentes aux comportements coopératifs, ces chercheurs ont trouvé que les gens collaborent parce qu'ils se sentent bien en le faisant. Dans une expérience sur le dilemme du prisonnier, le cerveau des sujets humains a été scanné au cours du jeu. Les résultats montrent que certaines zones du cerveau composées de neurones qui répondent à la dopamine (molécule qui joue un rôle dans le comportement relié à la dépendance) étaient fortement activées lors des séries de coopération mutuelle. Et cette réaction de plaisir neuronal était nettement moins élevée lorsque les participantes savaient qu'elles jouaient contre un ordinateur. Ainsi la perspective d'une alliance avec un autre être humain est source de plaisir. Selon les expérimentateurs, cette activation de neurones liés à la sensation de plaisir soutiendrait les relations sociales coopératives ; cette récompense pour une action coopérative serait un excellent

moyen d'inhiber les pulsions qui poussent à la défection. Toutefois, il n'est pas certain que ces résultats parlent réellement en faveur de la thèse de l'égoïsme psychologique. En effet, même si l'on a pu montrer que la récompense est liée aux actes coopératifs, il n'a pas été prouvé que l'anticipation de la récompense cause les choix coopératifs. Il se pourrait que la réaction neuronale soit simplement un effet secondaire, un épiphénomène des interactions coopératives sans influence d'une motivation sous-jacente (tout comme le plaisir de manger une pomme découle du fait de manger la pomme).

Le défenseur de l'égoïsme psychologique pourrait alors se référer à un récent article de Kevin HALEY et Daniel FESSLER (2005) qui montre que les gens sont largement influencés dans leurs choix coopératifs par certaines croyances intuitives liées aux rapports sociaux. Les expérimentateurs ont manipulé de manière subtile les paramètres du jeu en faisant apparaître des yeux stylisés sur les écrans d'ordinateur utilisés par une partie des sujets de l'expérience. Il se trouve que cette infime différence dans les conditions de jeu a largement influencé le taux de coopération des sujets. Il semblerait que les yeux stylisés soient perçus comme un indice de contrôle social qui incite à agir de manière pro-sociale par peur d'être puni. Ce genre de résultats remet en question les conditions d'expérimentation généralement utilisées pour tester les tendances comportementales des êtres humains en situation d'interaction sociale. Il se pourrait bien que les résultats de toutes ces expériences soient biaisés par des paramètres auxquels les expérimentateurs n'ont pas pensé. Par exemple il est probable que même sous condition d'anonymat total, les sujets ne parviennent pas à se défaire de l'impression d'être contrôlés si bien qu'ils sont poussés à agir de manière pro-sociale par crainte irraisonnée de la punition. Voilà qui ébranle un peu plus la thèse de l'altruisme psychologique. Cependant il ne s'agit ici que d'une hypothèse dont il n'est pas évident de prouver la pertinence. Personne ne niera que la pression du contrôle social pousse les gens à coopérer mais cela n'empêche pas que les expériences menées de manière précautionneuse (précisément celles qui évitent d'intégrer tout indice de contrôle social) permettent réellement de mettre en évidence des motivations de nature altruiste ou du moins pro-sociales chez les êtres humains (voir aussi FEHR & ROCKENBACH 2003).[177]

[177] Pour une controverse similaire relative à la motivation sous-jacente aux comportements réciproques indirects, voir LEIMAR et HAMMERSTEIN (2001) pour qui les individus agissent sous l'influence d'une

En fin de compte, pour ce qui est des motivations sous-jacentes à la tendance à coopérer, nous nous trouvons devant un match nul. Voyons ce qu'il en est de la punition altruiste ; relève-t-elle d'une motivation à rendre justice aux individus lésés ? Cette idée semble en réalité assez peu convaincante car il existe des interprétations concurrentes nettement plus plausibles. James FOWLER et ses collègues par exemple, pensent que la punition est causée par un sens de l'équité possédé par tous les êtres humains (2004 ; FOWLER *et al.* 2005). Ce sens de l'équité pousserait les sujets à punir les inégalités, c'est-à-dire les individus dont le gain est disproportionné par rapport à celui des autres. Une autre solution (d'ailleurs compatible avec la précédente) a été proposée par Ernst FEHR et Simon GÄCHTER (2004). Selon eux les êtres humains possèdent certaines normes sociales profondément ancrées dans leur esprit. La colère qui mène à la punition serait déclenchée lorsque les sujets prennent conscience que ces normes ne sont pas respectées. Une de ces normes pourrait être celle qui prône l'équité mais il en existe une autre qui paraît être encore plus puissante : il s'agit de la norme de réciprocité. L'idée est que si une personne s'engage dans la coopération, les autres doivent rendre la pareille (voir aussi BOWLES & GINTIS 2004). Si ces explications en termes de sens de l'équité ou norme de réciprocité sont correctes, alors on ne peut plus parler de motivation altruiste ; il s'agirait plutôt de motivation normative, au sens où c'est la prise de conscience d'une divergence entre la situation et la norme qui induit la motivation à punir.[178] A la rigueur, on pourrait même dire que la motivation provient d'une anticipation du plaisir de punir. C'est du moins ce que semble suggérer l'expérience suivante. Dominique DE QUERVAIN et collègues (2004) ont mené une expérience sur les réactions neuronales provoquées par la condition d'être victime d'un acte d'opportunisme. Dans le cadre d'un jeu de la confiance, les cerveaux des sujets auxquels on attribuait le rôle des premiers joueurs (les *trusters*) ont été scannés durant

---

motivation égoïste (maintenir une bonne réputation) ; pour la position opposée selon laquelle les individus sont motivés par une tendance à rechercher l'équité ou à punir les opportunistes, voir NOWAK & SIGMUND (1998) MILINSKI *et al.* (2001 ; 2002), WEDEKIND & MILINSKI (2000). MILINSKI et collègues rétorquent à LEIMAR et HAMMERSTEIN, que leur modèle théorique présuppose des capacités mentales que les êtres humains ne possèdent pas (mémoire prodigieuse, omniscience) et produisent des données empiriques qui semblent contredire la position de leurs contradicteurs (MILINSKI *et al.* 2001).

[178] Notons que cette interprétation infirme la thèse de l'égoïsme psychologique qui affirme que toute motivation est égoïste. On ne peut en revanche pas l'utiliser en faveur de la thèse de l'altruisme psychologique.

l'expérience. La focale était portée sur la partie du cerveau activée lorsque les sujets étaient victimes d'abus de confiance de la part des autres joueurs. En cas d'opportunisme de la part du deuxième joueur (c'est-à-dire s'il garde la totalité de la somme pour lui), le premier joueur avait la possibilité de punir soit de manière symbolique soit réellement ; dans le deuxième cas, la punition était coûteuse pour les deux partis. Les résultats montrent que lorsque le sujet choisit de punir réellement l'opportuniste, une zone subcorticale de son cerveau (appelée le « striatum ») est activée. Or des recherches antérieures ont montré que cette région du cerveau est activée lorsque l'on obtient une récompense et induit une expérience affective agréable. Il apparaît donc que les sujets obtiennent un sentiment de satisfaction lorsqu'ils punissent les opportunistes, c'est-à-dire lorsqu'ils prennent leur revanche. Mais ce n'est pas tout. La même expérience montre que le taux d'activation du striatum est corrélé avec le degré de punition ; plus le striatum est activé, plus grande est la somme investie par le punisseur pour se venger de l'opportuniste. Il semblerait donc que les sujets soient motivés à punir parce qu'ils anticipent une satisfaction due à la revanche. Plus l'anticipation est grande, plus ils sont disposés à payer de leur personne pour se venger.[179] Notons toutefois que cette expérience met en jeu des actes punitifs résultants du fait d'avoir été abusé par un autre joueur ; l'aspect de la vengeance entre donc en jeu. Il n'est pas certain que l'on obtienne des résultats similaires lorsque le punisseur est une tierce personne qui n'a pas elle-même été lésée (comme dans le cas de la variante du jeu de la confiance avec observateur).

Quoi qu'il en soit, il semblerait que la punition est loin d'être motivée par la prise en compte du bien-être et des intérêts d'autres joueurs. Mais cela n'invalide pas la thèse de l'altruisme psychologique. On peut tout à fait concéder que la motivation à la punition n'est pas de nature altruiste et ajouter que nous sommes motivés de manière altruiste pour réaliser d'autres actions.

En définitive, il semblerait que les données empiriques issues de la psychologie et de l'économie expérimentale ne peuvent être utilisées de manière convaincante ni en faveur de la thèse de l'altruisme psychologique ni en faveur de la thèse opposée (du moins dans l'état actuel de la recherche). Comme le fait justement remarquer Dale

[179] On trouve également cette ligne d'argumentation chez Michael PRICE, Leda COSMIDES et John TOOBY (2002).

JAMIESON (2002), il semblerait que l'on ne pourra jamais, sur la base d'approches expérimentales, être certain qu'une action particulière a été causée par une motivation altruiste ou égoïste.

Voyons s'il est possible de trouver un meilleur argumentaire chez les philosophes. Au cours de l'histoire de la philosophie, un bon nombre d'arguments ont été proposés pour contrer la thèse de l'égoïsme psychologique. Le plus fameux est dû à Joseph BUTLER (1991/1726, § 415) et porte sur la nature des désirs[180] qui motivent les sujets à agir de manière altruiste (ou du moins apparemment altruiste). BUTLER affirme que ces désirs ne sont pas de nature hédoniste, c'est-à-dire qu'ils ne sont pas dirigés vers notre propre plaisir. La prémisse de l'argument consiste à dire qu'un désir pour un objet extérieur doit être antérieur à la sensation de plaisir (laquelle découle de l'obtention de l'objet) ; dit autrement, une condition préalable pour éprouver du plaisir est d'avoir un désir orienté vers un objet. Par exemple, nous pouvons prendre du plaisir à manger une pomme uniquement si nous avons au préalable formé le désir de manger une pomme. Il s'ensuit que l'on ne peut pas dire avec l'égoïste psychologique que tout désir pour un objet est causé par un désir hédoniste préalable (en l'occurrence le désir du plaisir que cet objet est censé causer). Cet argument a fait école et on en retrouve diverses variantes dans les écrits de philosophes contemporains (BROAD 1930 ; FEINBERG 1984 ; NAGEL 1970).

Les tenants de la thèse de l'égoïsme psychologique pourraient toutefois rétorquer que l'hédonisme psychologique peut tout à fait s'accommoder de la prémisse de l'argument de BUTLER. Ils peuvent admettre que la condition nécessaire pour obtenir du plaisir est d'avoir eu au préalable un désir pour l'objet qui cause du plaisir. L'important est de savoir ce qui cause le désir pour l'objet et l'hédoniste pourrait affirmer qu'un désir pour un objet extérieur (par exemple une pomme) peut être suscité par le désir d'éprouver du plaisir (à supposer que l'on pense qu'il est plaisant de manger une pomme). En d'autres termes l'hédoniste peut accepter l'existence de désirs pour des objets externes[181] tout en affirmant que tous ces désirs sont eux-mêmes causés par un

[180] Nous verrons plus que loin que penser la controverse entre l'altruisme et l'égoïsme psychologique en termes de désir est fourvoyante.

[181] En principe l'hédoniste psychologique peut même admettre qu'une condition nécessaire pour avoir du plaisir est d'avoir eu au préalable un désir pour l'objet qui cause du plaisir. Notons cependant que cette affirmation peut être remise en cause pour d'autres raisons. Comme le fait remarquer Elliott SOBER, il y a des choses qui causent du plaisir sans que l'on ait eu un désir préalable pour l'objet lui-même ; l'odeur

désir hédoniste : celui d'obtenir du plaisir par le biais de l'obtention de l'objet. Nous aurions ainsi une chaîne causale du type

(a) Désir pour le plaisir → (b) Désir pour un objet extérieur → (c) Obtention de l'objet extérieur → (d) Plaisir

où (b) ne peut pas être produit sans (a).[182] D'autre part, pour se munir contre une série d'objections relatives à la plausibilité de sa position, le défenseur de l'égoïsme psychologique peut recourir à l'idée que (a) ne doit pas forcément être conscient. Cela lui permet d'affirmer par exemple que pour pouvoir être motivé à agir, le sujet doit se trouver dans un état d'inconfort (par exemple avoir faim ou être mal à l'aise à la vue de la souffrance d'autrui) qui lui fait concevoir un désir conscient ou inconscient de se débarrasser de cet état, c'est-à-dire un désir de type (a). Ou alors le sujet peut anticiper qu'en agissant d'une certaine manière (par exemple manger une pomme juteuse ou faire une bonne action), il obtiendra un sentiment agréable, et c'est (a) le désir conscient ou non d'éprouver le sentiment agréable qui le poussera à forger un désir de type (b) qui consiste par exemple à vouloir manger une pomme ou aider son voisin.[183]

Cet argument de l'inconscient est extrêmement puissant. Il permet de répondre à peu près toutes les tentatives philosophiques en faveur de l'altruisme psychologique. Considérons un autre exemple. Francis HUTCHESON a imaginé une expérience de pensée destinée à prouver l'existence de choix altruistes. « Supposons (…) que la Divinité déclare à un honnête homme qu'il va soudain l'anéantir mais, qu'à l'instant de sa mort, le choix lui soit laissé de rendre à l'avenir ses amis, ses enfants ou son pays heureux ou malheureux, alors qu'il ne pourra ressentir lui-même ni plaisir ni peine de leur état » (1991/1726, p. 152). Selon Hutcheson, dans de telles circonstances, la plupart d'entre

---

d'une fleur par exemple (1992 ; SOBER & D. WILSON 2003/1998, chap. 9 ; voir aussi PLATON, *Philèbe*, 51c-52b).

[182] Pour les détails de cet argument, voir SOBER 1992 ; SOBER & D. WILSON 2003/1998, chap. 9. Ces auteurs pensent que ni les expériences menées par les psychologues, ni les arguments philosophiques ne permettent de rejeter la thèse de l'égoïsme psychologique. Pour défendre la thèse de l'altruisme psychologique, ils proposent un argument évolutionnaire (2003/1998, chap. 10) qui sera mentionné plus loin (section 3.4.4).

[183] Une analyse phénoménologique de nos motivations par exemple montrerait que, par les moyens de l'introspection, nous ne pouvons souvent pas imaginer l'existence d'un désir de type (a) préalable à nos désirs de type (b).

nous choisirait la première option et seule l'existence de motifs altruistes permet d'expliquer un tel choix. L'élément clé de l'argument est le recours à l'introspection ; nous ne pouvons pas imaginer d'autres motifs qu'altruistes pour expliquer le choix de cet homme, ou le choix similaire que nous ferions si nous étions à sa place. Mais à nouveau, un partisan de l'égoïsme psychologique pourrait invoquer l'inconscient en rétorquant que l'introspection ne nous donne pas forcément accès à nos motifs les plus profonds. Il se pourrait bien que ces derniers consistent en un espoir irrationnel d'être récompensé dans l'au-delà ; dans ce cas, l'introspection nous tromperait sur nos réels motifs et la thèse égoïste demeure plausible.

Pour les partisans de l'égoïsme, le recours à l'inconscient comporte l'avantage d'immuniser leur interprétation contre les critiques (puisque, précisément, l'inconscient est insondable). Il y a cependant un revers à la médaille : l'inconscient implique que l'on ne peut rien prouver et il n'y a aucune raison de principe de favoriser la thèse de l'égoïsme psychologique ou la thèse opposée. En définitive, il semblerait que, sur le plan descriptif du moins, nous nous trouvions dans une impasse et que cette controverse demeurera à jamais irrésolue.

3.3.3. Une redéfinition des termes de la controverse

Pour sortir de cette impasse, je propose de reprendre le problème à la base en partant d'une distinction utile entre d'une part, la *motivation* altruiste et d'autre part les *motifs* altruistes (compris comme des buts, des désirs ou des intentions).

Suivant la tradition humienne, je pense que la *motivation* relève strictement de l'affect (cette thèse sera développée plus en détail à la section 5.2.2; voir aussi McShea & McShea 1999). Par exemple, une sensation de faim pousse l'agent à assouvir cette faim ; et cet état motivationnel dure jusqu'à ce que le tenaillement à l'estomac ait disparu. Quant aux *motifs*, ils sont généralement compris comme étant des désirs, des buts ou des intentions. Ils sont composés d'une partie affective et d'une partie explicitement cognitive ; ces deux aspects sont étroitement liés mais il est important de comprendre que si les motifs sont motivants, c'est uniquement en raison de leur composante affective. Prenons les désirs par exemple : le désir peut être compris comme le résultat de la combinaison d'un affect et d'une croyance. Plus précisément, il est

déclenché soit par une émotion (par exemple la pitié), soit par une simple sensation (par exemple la faim) et est dirigé vers un but fourni par une croyance (par exemple la croyance qu'il existe une pomme à proximité et que l'on peut s'en saisir et la manger). Le point important est que selon cette lecture, l'affect est la cause à la fois du désir et de l'action ; il a donc une priorité causale.

Cela posé, il me semble que la raison pour laquelle la controverse autour de l'altruisme a engendré autant de débats tient au fait que les auteurs réfléchissent en termes d'intention, de buts ou de désirs. Mais si l'on questionne la possibilité de l'existence d'actions altruistes, la focale doit être mise sur la *motivation* et non sur les *motifs* car un motif n'est motivant qu'en vertu de sa composante affective. En bref, puisque la controverse concerne la manière dont les actions ont été causées, l'aspect motivationnel a priorité.

Ainsi, il me paraît utile de faire une distinction entre un *altruisme psychologique motivationnel*, qui est une forme de réaction émotionnelle (à ce propos, voir section 5.2.1.iii), et un *altruisme psychologique sophistiqué* qui relève de ce que les gens considèrent comme étant les motifs de leurs actions (nous verrons aux sections 4.2.2 et 5.4 que la moralité est intrinsèquement liée à cette seconde forme d'altruisme).

Mettre en place cette distinction entre motivation (altruisme psychologique motivationnel) et motif (altruisme psychologique sophistiqué) signifie reformuler la controverse. Puisque cette dernière porte sur l'existence d'actions altruistes psychologiques « en vertu de leur causes », la question pertinente à se poser est de savoir s'il existe des *motivations* altruistes capables de produire ce genre d'actions. Et d'autres termes, il s'agit de sonder la possibilité d'affects altruistes capables de nous motiver à agir.

En s'inspirant de la définition de l'altruisme psychologique proposée plus haut, on peut dire qu'un affect est altruiste s'il porte sur le bien-être et les intérêts d'autrui. Selon la même logique, pour qu'un affect puisse être considéré comme égoïste, il faut qu'il porte sur nos propres intérêts et bien-être. Cela implique qu'une condition nécessaire pour que des affects puissent être traités comme altruistes ou égoïstes est qu'ils possèdent un contenu intentionel (au sens où ils portent sur un objet du monde extérieur). Or ce n'est pas le cas de tous les affects. Par exemple les simples sensations comme une douleur ressentie après avoir été frappé ou le tenaillement de la faim sont des affects qui ne sont dirigés vers aucun objet ; ils sont non intentionels. On pourrait

ici, affirmer d'emblée que la thèse de l'égoïsme psychologique est fausse puisque des affects non intentionels comme un tenaillement à l'estomac peuvent motiver à agir. Mais ce ne serait peut-être pas faire justice à la controverse, laquelle porte sur des motifs/motivations qui ont pour objet le bien-être et les intérêts d'individus. Concentrons-nous donc sur la question de l'existence d'affects intentionels en vertu du fait qu'ils sont dirigés vers le bien-être d'individus.

Lorsqu'un affect dépasse le simple état de sensation et est dirigé vers un objet, il devient une émotion comprise au sens défendu par des philosophes comme Peter GOLDIE (2000) ou Sabine DÖRING (2007). Selon ces auteurs, une expérience émotionnelle est une « sensation envers » (*feeling towards*), un état affectif intentionel. Ainsi, la motivation altruiste, si elle existe, est une émotion ; chaque fois que nous sommes motivés de manière altruiste (si cet état existe), nous faisons une expérience émotionnelle. Nous pouvons donc cerner encore mieux la controverse : il s'agit de savoir s'il existe des émotions qui portent sur le bien-être et les intérêts d'autrui[184] *et* sont capables de nous motiver à agir.

Une première constatation est que parmi les émotions, il y a celles qui sont dirigées vers soi-même (par exemple la honte, la fierté), il y a celles qui sont dirigées vers autrui (par exemple l'amour, l'admiration) et il y a celles dont on ne peut pas vraiment dire qu'elles soient dirigées vers quelqu'un (par exemple la peur qui est dirigée vers ce qui est dangereux, ou la joie). Parmi les émotions dirigées vers autrui, il y en a clairement qui sont relatives au bien-être et aux intérêts d'autrui. Comme exemples d'émotions altruistes, on peut compter l'amour, la sympathie, la compassion, certaines formes d'admiration ou de respect (par exemple lorsque l'on est témoin d'une action qui améliore le sort d'une personne). Toutes ces émotions montrent que nous ne sommes pas indifférents au bien-être et aux intérêts d'autrui ; c'est donc celles-là qui sont au cœur de l'*altruisme psychologique motivationnel*.

D'autre part, le phénomène émotionnel peut être compris comme un modèle tripartite (pour plus de précisions, voir la section 5.3.1). Certaines causes typiques causent certaines réactions physiologiques et cognitives typiques lesquelles sont liées à des tendances typiques à l'action. Par exemple un épisode de compassion est une

[184] En réalité, pour être précis, il faudrait encore ajouter une distinction supplémentaire pour éviter de devoir compter dans le camp des altruistes des émotions comme ce que l'on appelle en allemand *Schadenfreude*, qui consiste à prendre du plaisir à la vue du malheur d'autrui.

réaction à un ensemble de circonstances comme le fait d'être témoin de la souffrance d'autrui. Cet épisode de compassion se caractérise par un certain état du système neuronal et endocrinien ainsi que par une expression faciale spécifique de la compassion. Enfin, un sujet compatissant sera motivé à venir en aide à l'individu qui souffre. Vu sous cet angle, un phénomène émotionnel altruiste doit être conçu comme un ensemble d'événements intimement liés les uns aux autres ; et une réaction émotionnelle révèle un mécanisme psychologique qui se met en branle face à certains *inputs* et produit certains *outputs*. Il me semble que réfléchir de cette manière permet de faire taire un certain nombre d'interprétations égoïstes de notre motivation à agir. Je m'explique. Il se peut qu'en décomposant un mécanisme émotionnel altruiste comme celui de la compassion, il apparaisse que l'état du système neuronal et endocrinien correspond à un affect négatif qui induit le sujet à accomplir des actions qui vont avoir pour conséquence de le libérer de cet état. Mais cela ne peut pas être interprété en termes de motivation égoïste car ce qui importe est de déterminer si le mécanisme dans son ensemble est altruiste, c'est-à-dire s'il a été déclenché par la prise en comte du bien-être et des intérêts d'autrui (quels que soient les processus physiques et endocriniens impliqués dans la mise en action du mécanisme en question). Le fait que la compassion s'exprime par une phénoménologie déplaisante et disparaît avec l'input qui l'a causée (c'est-à-dire lorsque les besoins d'autrui sont satisfaits) n'en fait pas une émotion égoïste.

La dernière inconnue à résoudre concerne la question de savoir si les émotions altruistes sont capables de nous motiver à agir ou si elles sont toujours contrebalancées par d'autres émotions non altruistes. Une réponse positive nous permettra d'affirmer que la motivation altruiste existe et peut orienter nos actions, donc que la thèse de l'altruisme psychologique sort vainqueur de la controverse.

Il nous faut des arguments ici car on pourrait imaginer que les êtres humains sont toujours sujets à des conflits d'influence entre différents mécanismes motivationnels et qu'en fin de compte, les mécanismes altruistes sont si faibles qu'ils ne prennent jamais le dessus. Par exemple si l'on reprend le phénomène présenté plus haut de l'observateur punisseur dans le jeu de la confiance (p. 147), on pourrait imaginer que deux mécanismes motivationnels entrent en jeu. D'un côté le mécanisme de la compassion inciterait l'observateur à punir l'opportuniste afin d'aider le joueur lésé. D'un autre côté l'observateur serait ébranlé par un mécanisme émotionnel de réaction normative qui n'est ni altruiste, ni égoïste : il serait outré de voir que l'un des joueurs contrevient à la

norme de réciprocité (ou à celle d'équité) et cela l'inciterait à le punir.[185] Si la poussée motivationnelle du premier mécanisme est infime par rapport à celle du second, alors il sera difficile de considérer l'action de l'observateur comme altruiste. Le problème dans cet exemple est que les deux mécanismes émotionnels poussent à la même action si bien qu'il n'y a pas moyen de définir si le second mécanisme est seul responsable de l'action.

Dans ce qui suit, je vais utiliser deux lignes d'argumentation pour montrer que les émotions altruistes exercent un impact non négligeable sur nos choix d'action. La première (développée dans la section suivante) consiste à donner la parole aux psychologues qui, je pense, ont montré de manière convaincante à quel point les émotions altruistes exercent un effet incitatif sur notre comportement pro-social. La seconde ligne d'argumentation (section 3.4), moins directe, consistera à explorer les raisons évolutionnaires de l'apparition des motivations altruistes.

3.3.4. La psychologie en faveur de l'altruisme

La psychologue Nacy EISENBERG (1986 ; 2006) a récolté une quantité impressionnante de données montrant que les émotions altruistes ont pour effet de contraindre de manière significative les choix purement égoïstes et de promouvoir les comportements pro-sociaux. Il apparaît notamment qu'au fur et à mesure que les enfants grandissent, les motivations de leurs actions pro-sociales sont moins liées aux récompenses et dépendent plus de la prise en compte des intérêts d'autrui.[186]

Dans la même ligne, le psychologue Daniel BATSON (1991) a mis en place toute une batterie de tests psychologiques pour montrer que l'émotion de sympathie[187] incite les sujets à venir en aide à des personnes en difficulté. Sous différentes variantes, le

[185] Même s'il est nettement moins probable que ce soit le cas, on pourrait même imaginer une motivation égoïste, une sorte d'émotion d'excitation positive qui consiste à anticiper le plaisir causé par le fait de punir l'opportuniste.

[186] « It appears that the cognition associated with children's prosocial actions becomes more internal and less related to external gain with development. » (EISENBERG 1986, p. 92)

[187] En réalité BATSON utilise la notion d'« empathie » mais il la définit, non comme une simple capacité cognitive de comprendre les émotions d'autrui (définition à laquelle je donne ma préférence), mais comme une émotion dirigée vers autrui qui inclut des sentiments de sympathie ou de compassion. (BATSON 1991, pp. 86-87)

principe de l'expérience consistait à mettre un sujet dans une salle contenant un écran de télévision qui reproduisait en direct ce qui se passait dans une autre salle où une prétendue étudiante (en réalité une excellente actrice) recevait des chocs électriques. On donnait alors la possibilité au sujet de remplacer l'étudiante et subir les chocs à sa place. Au cours de ces tests, BATSON a fait principalement varier deux paramètres : le degré de sympathie que les sujets ressentent envers l'étudiante[188] et la possibilité (facilitée ou non) pour les sujets de quitter l'expérience en cours de route. L'analyse de la probabilité avec laquelle les sujets étaient prêts à remplacer l'étudiante en fonction de la variation des paramètres mentionnées ci-dessus permet de remettre sérieusement en question un bon nombre d'hypothèses égoïstes sur la motivation à l'action pro-sociale ; notamment l'idée que les gens aident pour ne pas essuyer des reproches de la part d'autrui, pour ne pas se sentir coupables de ne pas l'avoir fait , pour cesser de se sentir mal à l'aise face à la souffrance ou au malheur d'autrui (CIALDINI *et al.* 1987), pour recevoir les louanges et les honneurs, pour se sentir bien après avoir fait une bonne action (pour une excellente revue de ces débats, voir SOBER & D. WILSON 2003/1998, pp. 260-274).

Au vu de ces résultats[189], il semblerait que les émotions altruistes exercent effectivement un effet bénéfique sur notre comportement pro-social.

3.4. L'évolution de l'altruisme psychologique

Avec beaucoup d'autres auteurs (notamment JAMIESON 2002 ; KITCHER 1998), je pense qu'il y a de bonnes raisons évolutionnaires de croire en l'existence de motivations altruistes suffisamment fortes pour avoir une répercussion sur l'action des sujets. Si les émotions altruistes ont pu évoluer, c'est très probablement parce qu'elles exercent un effet sur le comportement des gens.[190] Il vaut donc la peine de se tourner vers les théories évolutionnistes et passer en revue les explications de l'évolution de l'altruisme psychologique qu'elles proposent. Une explication évolutionnaire convaincante de

[188] Pour manipuler le degré de sympathie, on fournissait au sujet des informations sur l'étudiante qui correspondaient ou non à la manière dont le sujet s'était décrit lui-même en remplissant un questionnaire avant le début du test.

[189] Pour d'autres données de ce type issues de la psychologie et de la sociologie, voir PILIAVIN & CHARNG 1990.

[190] A moins qu'il ne s'agisse d'un effet dérivé mais nous verrons plus loin (section 3.4.6) que cette hypothèse est peu convaincante.

l'altruisme psychologique pourrait fournirait un argument supplémentaire contre la thèse de l'égoïsme psychologique. Nous verrons que du même coup, cela permettra de mieux saisir les rapports entre l'altruisme évolutionnaire et l'altruisme psychologique.

Commençons par une distinction utile. Lorsque l'on cherche à expliquer des phénomènes d'un point de vue évolutionnaire, il est très utile de réfléchir en termes de *causes proximales* et de *causes ultimes*.[191] Les causes proximales sont celles qui agissent du vivant de l'organisme. En revanche, les causes ultimes renvoient à un temps antérieur à la vie de l'organisme : au temps de l'histoire de l'espèce. Considérons l'exemple d'un homme qui sauve un enfant de la noyade. Une explication en termes proximaux dirait que l'homme a fait cette action parce que la détresse de l'enfant l'a touché ; la compassion qu'il a ressentie à l'égard de cet enfant l'a poussé à se jeter à l'eau pour le sauver. Une explication en termes ultimes par contre pourrait être celle-ci : l'homme a sauvé l'enfant parce qu'au cours de l'évolution génétique et culturelle des êtres humains de la région dans laquelle il a grandi, les individus ont acquis la capacité de ressentir de la compassion face à la détresse d'autrui ; ce mécanisme psychologique qui incite à aider autrui a évolué parce qu'il s'avère évolutionnairement avantageux pour l'espèce humaine.

Lorsqu'il est question de déterminer les motivations particulières qui ont causé l'action d'un agent, il faut penser en termes de causes proximales tout en gardant à l'esprit qu'à l'échelle de l'évolution, les mécanismes sous-jacents aux motivations particulières peuvent trouver une explication en termes d'avantages adaptatifs. Pour ce qui est de l'altruisme psychologique, nous avons vu que les mécanismes proximaux sont les émotions altruistes. Voyons donc comment ces émotions se traduisent en termes de *fitness* ; sont-elles avantageuses pour les gènes qui induisent ces émotions, pour l'agent lui-même ou pour le groupe duquel il fait partie ? Voyons aussi quelles sont les conditions de leur apparition et de leur stabilisation au cours de l'évolution. Dans ce qui suit, je vais passer en revue un certain nombre de théories, plus ou moins compatibles entre elles, qui ont été formulées à cet effet.

[191] Cette distinction est due à Ernst MAYR (1961).

3.4.1. Les bases posées par la sélection de parentèle et la réciprocité

Nous avons vu que la théorie de la sélection de parentèle développée par William HAMILTON permet d'expliquer la raison pour laquelle, chez les espèces dont les comportements sont entièrement régis par l'instinct, les comportements altruistes *évolutionnaires* envers des proches parents ont pu être sélectionnés au cours de l'évolution ; l'explication réside dans la propagation des gènes qui codent pour ces comportements (voir section 2.2.1.ii). Avec bien d'autres auteurs, je pense que cette même explication peut s'étendre aux causes proximales qui causent ces comportements ; les gènes peuvent régir (du moins en partie) des mécanismes émotionnels comme la sympathie ou la compassion (SOBER & D. WILSON 2003/1998 ; TOOBY & COSMIDES 1989 ; RICHARDS 1993). Richard JOYCE (2006, pp. 47-48) exprime très bien ce phénomène avec l'exemple de l'amour :

> « La sélection naturelle cherche à inciter les gens à agir pour le bien de leurs gènes ; pour y parvenir elle utilise, comme mécanisme prochain, l'émotion de l'amour. (…) Lorsque certains des gènes cruciaux résident dans d'autres individus (en particuliers nos proches parents), la solution naturelle est de créer un amour (discriminant et conditionnel) dirigé directement vers autrui. » (JOYCE 2006 p. 48, ma traduction)[192]

Dit autrement, si la *fitness* reproductive a été améliorée par le fait d'aider les membres de sa famille, le processus de la sélection naturelle pourrait bien avoir développé nos cerveaux de manière à ce que l'on ressente de l'amour envers des membres de notre famille. JOYCE ajoute à juste titre que même si cet amour est strictement dirigé vers les proches parents, il n'en demeure pas moins qu'il est sincère, non instrumental et profondément dirigé vers autrui. En effet, il serait ridicule de dire que les parents aiment leurs enfants de manière égoïste parce qu'ils ont des motivations inconscientes concernant leur *fitness* inclusive. Affirmer cela reviendrait à confondre les causes proximales et les causes ultimes.

[192] « Natural selection wants to get people acting for the sake of their genes, and employs the emotion of love as a proximate mechanism to achieve this. (…) When some of the genes that matter reside in other individuals (one's kin, in particular), the natural solution is to create a (discriminating and conditional) non-derivative, other-directed love. » (JOYCE 2006 p.48)

Outre la sélection de parentèle, il semblerait que les mécanismes liés à la réciprocité apportent également quelques lumières sur l'évolution de l'altruisme psychologique. D'après Robert TRIVERS (1985, 1971), le modèle de la sélection de l'altruisme réciproque permet non seulement d'expliquer l'évolution de comportements apparemment altruistes évolutionnaires, mais également les systèmes psychologiques sous-jacents à ces comportements. Selon lui, l'altruisme réciproque a joué un rôle important dans l'évolution humaine, à tel point qu'un bon nombre d'émotions sont apparues et se sont stabilisées au fil de l'évolution parce qu'elles avaient pour fonction biologique de favoriser le système de réciprocité. Voici un aperçu de la manière dont les choses ont dû se passer. Pour commencer, il y a eu l'évolution d'un système simple de réciprocité soutenu par quelques tendances purement instinctives. Du point de vue de l'évolution, la réciprocité est extrêmement avantageuse et crée un environnement social dans lequel il vaut la peine d'agir de manière pro-sociale. Mais en même temps, dans un monde de coopérateurs, il est tentant de tricher. Ainsi, s'il y a sélection de comportements altruistes, parallèlement, il y a sélection de comportements opportunistes. Et s'il y a opportunisme, il y a une sélection des moyens de détecter et de contrer l'opportunisme, lequel évoluera vers des formes de plus en plus subtiles, et ainsi de suite. Dans le cadre d'une telle dynamique, toute une batterie de mécanismes et de capacités a pu être sélectionnée. Pour ce qui est des émotions,[193] la sympathie a été sélectionnée pour motiver à agir de manière coopérative et à aider de futurs partenaires de coopération dans le besoin ; la gratitude a été sélectionnée parce qu'elle manifeste une sensibilité au rapport coût-bénéfice de l'action altruiste et incite le bénéficiaire à rendre la pareille ; la culpabilité a été sélectionnée parce qu'elle décourage les opportunistes et les pousse au repentir et à des actions réparatrices ; enfin, l'indignation qui pousse à des actes punitifs a été sélectionnée parce qu'elle protège les altruistes contre les opportunistes.[194] Cette liste n'est pas close ; on pourrait par exemple ajouter l'émotion d'amitié pour son effet bénéfique sur le renforcement des liens de réciprocité,

[193] Evidemment, en plus des mécanismes et capacités favorisant la pro-socialité, on constate une évolution parallèle de capacités et méthodes de tricherie toujours plus subtiles ainsi qu'un raffinement des capacités de détecter les tricheurs. Je n'en parlerai pas ici puisque l'objet de cette section est de définir les causes de l'évolution des émotions altruistes.

[194] Il va sans dire que malgré leur caractère inné, ces émotions altruistes n'en demeurent pas moins extrêmement plastiques ; elles peuvent être éduquées et adaptées aux conditions locales.

etc.

En résumé, il est fort probable que certaines émotions altruistes aient évolué parce qu'il s'agit de mécanismes adaptés à l'environnement social complexe décrit plus haut et parce qu'elles permettent de renforcer et réguler le système de la réciprocité. Toutefois, ce scénario a ses limites : il implique que nous dirigions nos émotions altruistes de manière sélective, envers des personnes que nous côtoyons régulièrement (ou au mieux envers des individus qui nous ressemblent ou qui ressemblent à ceux que nous côtoyons régulièrement).

Toutefois, il est bien connu que nous sommes capables d'éprouver de la compassion, de la pitié, de l'amour, etc. envers des inconnus. Mère Teresa était certainement une personne remplie de sentiments altruistes envers les miséreux. Ainsi, les deux explications proposées ne suffisent pas à rendre compte de tous les ressorts de l'évolution de sentiments altruistes puisqu'elles impliquent que nous dirigions nos sentiments de manière discriminatoire, soit en faveur de nos proches parents, soit en faveur de potentiels partenaires de réciprocité. La réponse à l'évolution de l'altruisme psychologique fournie par ces deux théories ne peut donc être que partielle. Notons que je ne voudrais pas en minimiser l'apport. Je pense avec Richard JOYCE (2006, p. 49) qu'elles sont suffisantes pour expliquer les raisons évolutionnaires de la mise en place des mécanismes neuronaux nécessaires à la motivation altruiste. Il s'agit maintenant de comprendre comment il se fait que nous ayons étendu le champ des objets sur lesquels portent ces émotions ; comment par exemple, nous en sommes venus à être compatissants devant des inconnus en détresse ou à aimer non seulement nos enfants mais également toute sorte d'autres personnes. Dans les sections suivantes, je vais présenter une suite d'hypothèses plus ou moins compatibles entre elles, mais dont certaines sont plus convaincantes que d'autres.

3.4.2. La théorie du vestige

La théorie du vestige (*vestige theory*) a déjà été mentionnée à la section 2.3.6. Ses défenseurs les plus connus sont les psychologues évolutionnistes John TOOBY et Leda COSMIDES (1989). Selon eux, les émotions altruistes (comme la sympathie ou l'amitié) seraient une relique de notre passé. On trouve ce trait aujourd'hui parce que les gènes qui en sont responsables se sont implantés dans le pool génétique de l'espèce humaine

au cours du pléistocène. En ce temps-là, les individus vivaient dans de petits groupes majoritairement constitués de proches parents si bien qu'ils ne rencontraient pas souvent d'étrangers. Grâce à cet environnement particulier et à la force de la sélection de parentèle, les tendances altruistes psychologiques non discriminatoires ont pu être sélectionnées. Mais entre temps, les conditions environnementales ont changé (nous ne vivons plus dans de petits groupes d'individus apparentés) si bien que de nos jours, ce trait altruiste psychologique n'est plus adapté ; les émotions altruistes poussent les individus à dépenser leur énergie envers des individus non apparentés, c'est-à-dire des individus qui ont une plus faible probabilité de posséder les gènes qui induisent ces émotions.[195]

Nous avons vu à la section 2.3.6 que l'analyse de TOOBY et COSMIDES laissait à désirer ; si les groupes étaient effectivement composés en grande majorité de proches parents et qu'il y avait peu de migration et peu de contacts entre les différents groupes, alors l'effet de la sélection de parentèle devait probablement être nul.

Il y a peut-être moyen de donner du crédit à la théorie du vestige en échafaudant une explication similaire qui ne fasse pas appel au principe de sélection de parentèle mais plutôt à celui de la réciprocité (laquelle apporte un avantage au niveau individuel). Voici comment elle pourrait fonctionner. Dans les petits groupes du pléistocène, les individus se connaissaient et interagissaient régulièrement si bien que tout le monde avait intérêt à pratiquer la réciprocité. Dans un tel climat de réciprocité, les émotions altruistes non discriminatoires ont pu évoluer pour faciliter les relations de réciprocité. Mais de nos jours, le contrôle social et les interactions répétées sont plus difficiles à réaliser puisque nous vivons dans de très grands groupes. En conséquence, les émotions altruistes s'avèrent être des traits désavantageux pour les individus qui les portent.

En fait, cette explication n'est pas plus convaincante car elle oublie un principe qui a déjà été mis en évidence par AXELROD (1996/1984, p. 129) et John MAYNARD SMITH (section 2.2.2.iv): dans un monde de coopérateurs inconditionnels, il suffit d'un tricheur pour déséquilibrer tout le système de coopération.

[195] Selon la même logique, on peut ajouter que l'altruisme psychologique sera probablement désélectionné au cours de l'évolution future (à moins qu'il ne soit intrinsèquement lié à d'autres capacités évolutionnairement avantageuses).

3.4.3. La théorie du signal coûteux

Selon la théorie du signal coûteux, les émotions altruistes seraient d'excellents signaux indicateurs des motivations des individus. En effet, il est difficile de fausser ou d'imiter une émotion si bien que lorsqu'une personne est émue par la souffrance d'autrui, elle signale par la même occasion sa propension à agir de manière pro-sociale. Ce signal est à la fois un handicap et un avantage. Le handicap tient à ce que l'individu altruiste pourra plus facilement être la proie des opportunistes. L'avantage est qu'une personne altruiste sera respectée dans la société et bénéficiera d'une réputation de bon partenaire d'interactions sociales, si bien que beaucoup d'individus chercheront à engager des relations de réciprocité avec elle. Selon Robert FRANK (1988), Richard ALEXANDER (1979 ; 1987 ; 1993) ou Amotz ZAHAVI (1977, 2002, p. 253), le fait de posséder des tendances altruistes profondément ancrées en nous et qui se manifesteraient publiquement à notre insu apporterait un avantage au niveau individuel.

Cette théorie, qui reste tout de même très spéculative, n'est probablement pas suffisante à elle seule pour expliquer l'évolution des émotions altruistes. Elle pourra être complétée par celles qui vont être présentées dans les deux sections suivantes.

3.4.4. La théorie du moyen heuristique le plus efficace

La théorie du moyen heuristique le plus efficace repose sur le principe que dans certains environnements sociaux, le coût évolutionnaire de l'égoïsme est tellement élevé qu'il vaut la peine d'être altruiste. On trouve cette idée chez un bon nombre d'auteurs ; certains se contentent de la stipuler (GIBBARD 2002/1990, p. 102), d'autres la développent en détails en y ajoutant chacun une touche particulière en fonction du système global qu'ils défendent. De manière plus ou moins explicite, par exemple, les anthropologues évolutionnistes (RICHERSON, BOYD & HENRICH 2003 ; BOWLES & GINTIS 2002) présentent la punition altruiste (section 2.3.4) comme un facteur clé de l'évolution de l'altruisme psychologique. Cette théorie part du principe qu'à un moment de l'histoire humaine, les normes sociales renforcées par la punition (punition altruiste) ont émergé. La présence de ces normes renforcées a créé un environnement social dans lequel l'opportunisme n'est pas une stratégie qui vaut la peine d'être pratiquée, car la désobéissance aux normes sociales est durement sanctionnée. Dans un environnement social tel que celui-ci, il est probable que des dispositions psychologiques altruistes

s'avèrent avantageuses du point de vue de la sélection naturelle. Voici les détails de l'argument. Un individu égoïste psychologique calcule ses intérêts et développe un comportement opportuniste chaque fois qu'il pense pouvoir éviter la sanction ; parfois il se trompe, ce qui entraîne une sanction et parfois son opportunisme porte ses fruits au détriment des individus coopérateurs. Un individu altruiste psychologique par contre, ne prend pas le temps de calculer ses intérêts et agit invariablement de manière coopérative ; parfois, il doit payer le coût des actions opportunistes des égoïstes psychologiques, en revanche il ne perd pas de temps et d'énergie en calcul d'intérêts et ne risque jamais d'être puni. Considérons les coûts engendrés par l'utilisation de chacune de ces deux stratégies. Dans le cadre d'un environnement social à la fois complexe et comprenant des punisseurs altruistes, les égoïstes psychologiques se voient imposer une dépense d'énergie considérable en calculs pour décider à quel moment il vaut la peine d'agir de manière opportuniste. D'autre part, cette même complexité de l'environnement social augmente les risques de mauvais calculs suivis d'une sanction. L'altruiste psychologique en revanche ne souffre pas de la complexité de l'environnement social ; en coopérant sans réfléchir, il s'épargne à la fois les risques de la punition et de grands efforts cognitifs. Or, ces deux types d'avantages sélectifs en faveur des altruistes psychologiques pourraient bien dépasser les avantages occasionnels obtenus par les égoïstes psychologiques. C'est du moins ce que pensent les anthropologues évolutionnistes.

Ainsi, le fait de posséder des dispositions psychologiques altruistes (c'est-à-dire des émotions altruistes) peut, dans certaines circonstances sociales, s'avérer avantageux du point de vue de la sélection naturelle. L'altruisme psychologique peut évoluer parce qu'il est un moyen heuristique efficace pour naviguer dans des environnements sociaux complexes composés de normes renforcées par la punition. Ou bien dit plus crûment : les dispositions altruistes psychologiques ont évolué parce qu'elles ont servi à éviter la punition. Il est intéressant de noter ici que l'on se trouve en présence d'un cas de coévolution gène-culture, ou plus précisément, il s'agit d'un exemple d'effet Baldwin (section 1.2.3, p. 45) : sur le long terme, un phénomène culturel exerce un impact sur l'évolution génétique.

Philip KITCHER (2006) a développé une théorie assez similaire. Il pense que

l'altruisme psychologique[196] a évolué parce qu'il s'agit d'une stratégie adaptative dans un contexte de jeux de coalitions (*coalition games*). Plus précisément, son explication fonctionne de la manière suivante. Etant donné la complexité des interactions à l'intérieur de coalitions et étant donné le prix d'une rupture sociale (perte de temps et d'énergie à rétablir la paix), il n'est pas avantageux de procéder à des calculs d'intérêts pour décider si l'on veut coopérer ou non à l'intérieur d'une coalition ; outre son côté risqué, cette activité serait trop coûteuse en termes de temps et d'énergie. Dans un tel contexte, une tendance psychologique à répondre aux préférences d'autrui, si elle est suffisamment représentée dans la population, favorise le développement de coalitions saines étendues et productives, avec tous les avantages qu'elles apportent.

Dans la même veine, Elliott SOBER et David WILSON (2003/1998, chap. 10) comparent la plausibilité de l'évolution de mécanismes propres à l'égoïsme motivationnel par rapport à l'évolution de mécanismes propres au pluralisme motivationnel (ce dernier implique à la fois des motivations égoïstes et altruistes). Pour les besoins de l'argument, ils se concentrent sur le cas des soins parentaux chez les êtres humains et se demandent si, pour expliquer ce phénomène, il est plus probable que l'égoïsme ou le pluralisme motivationnel ait été sélectionné. Au terme d'une analyse assez complexe dont je ne présenterai pas les détails ici, ils concluent que le second mécanisme est le plus probablement responsable des comportements de soins parentaux ; tout en faisant l'économie des calculs d'intérêts, il s'avère plus fiable et réalise mieux sa fonction biologique qui est d'assurer la survie de la progéniture.

Assurément ces approches peuvent être critiquées dans le détail mais il me semble qu'elles comportent toutes une part de vérité : dans certains environnements sociaux, il vaut la peine d'être altruiste psychologique plutôt qu'égoïste.

3.4.5. La théorie de la sélection culturelle et l'effet Baldwin

Susan BLACKMORE (1999) part du principe que la sélection génétique à elle seule ne peut pas engendrer de l'altruisme psychologique au-delà du cercle familial et de la

[196] Notons que KITCHER (2006, p. 164) conçoit l'altruisme psychologique non comme une émotion mais comme une tendance aveugle à répondre aux préférences d'un autre individu avec lequel on est engagé dans une activité coopérative. Le modèle qu'il propose reste toutefois applicable avec une compréhension de l'altruisme psychologique en termes d'émotions.

réciprocité. Pour expliquer l'évolution de l'altruisme envers des inconnus, il faut selon elle recourir au mécanisme de sélection culturelle de groupe et à l'effet Baldwin (section 1.2.3, p. 45). Selon elle, au départ, l'altruisme élargi ne peut être qu'un phénomène culturel qui s'apprend et se transmet d'un individu à l'autre par le biais de l'apprentissage et de l'imitation.[197] Dans certains environnements, les entités culturelles altruistes (les « mèmes » altruistes), qui consistent par exemple en des actions, attitudes ou principes pro-sociaux, peuvent être transmises de manière assez efficace parce qu'elles assurent une reconnaissance sociale significative aux personnes qui les expriment (si bien qu'elles sont fréquemment imitées et enseignées au sein du groupe). Ce phénomène a pour effet de renforcer la cohésion et la force du groupe, et au fil de l'évolution culturelle, ces groupes seront sélectionnés au détriment des groupes dans lesquels l'altruisme n'est pas pratiqué ; c'est ainsi que les mèmes altruistes peuvent envahir toute une population. Après un grand nombre de générations reproduisant et propageant les mèmes altruistes, grâce à l'effet Baldwin, de véritables mécanismes psychologiques altruistes ont pu être ancrés dans notre matériel génétique.

Cette théorie combine à la fois celle de la sélection culturelle de groupe et celle du signal coûteux. Si l'on en retire l'idée spécieuse de « mème altruiste » (pour une critique, voir section 1.2.2) et que l'on se contente de dire que les individus apprennent à se comporter de manière altruiste, alors elle peut s'avérer parfaitement crédible.

Il est probable que les différents mécanismes proposés dans les trois dernières sections ont contribué de manière conjointe à l'évolution de l'altruisme psychologique élargi. Pour terminer, je vais présenter et exprimer quelques doutes sur une dernière théorie ; celle du produit dérivé.

3.4.6. La théorie du produit dérivé

Selon la théorie du produit dérivé, les mécanismes sous-jacents à l'altruisme psychologique ne sont pas un produit direct de l'évolution ; ils n'ont pas été sélectionnés parce qu'ils favorisaient la transmission de gènes ou parce qu'ils étaient

[197] James BALDWIN (1980/1909) lui-même a maintenu que la transmission des impulsions altruistes procède au moyen de l'imitation, de l'éducation et de l'apprentissage. De même, selon lui, cette transmission est favorisée par la sélection de groupe.

favorables aux individus qui possédaient ses mécanismes ou parce qu'ils étaient avantageux pour le groupe dans lequel les individus porteurs évoluaient. Ils sont simplement étroitement liés à un ou plusieurs autres traits qui eux sont adaptatifs.[198]

Ronald DE SOUSA mentionne (sans la soutenir explicitement) une solution de ce type (2001, p. 116)[199]. La sympathie, émotion altruiste par excellence, pourrait être un effet dérivé de la capacité de lire dans l'esprit des gens, c'est-à-dire la théorie de l'esprit.[200] Voici comment cela fonctionne dans le détail :

La théorie de l'esprit (voir aussi p. 137) permet de lire les états d'esprit d'autrui (ou du moins de se les représenter). Elle comporte l'indéniable avantage de prédire avec un bon taux de réussite le comportement d'autrui, ce qui permet de décider de son propre comportement en fonction de cette prédiction. Du point de vue stratégique, le fait de posséder la théorie de l'esprit est extrêmement avantageux et il paraît clair que c'est la raison pour laquelle elle a été sélectionnée.[201]

Il semblerait que le mécanisme de la théorie de l'esprit soit indissolublement lié à celui de l'empathie cognitive, qui est la capacité de comprendre ou saisir ce que ressent autrui, à lire (ou du moins d'imaginer) les états émotionnels d'autrui.[202] Les deux

---

[198] A ce propos, voir la distinction entre sélection *pour* et sélection *de* proposée par Elliott SOBER (p. 26).

[199] Dans le même ordre d'idée, selon Stephen GOULD et Richard LEWONTIN (1979), l'altruisme est un effet dérivé non adaptatif qui surferait sur d'autres traits adaptatifs comme l'intelligence, la théorie de l'esprit, la conscience de soi ou l'empathie. Ces capacités combinées produiraient l'altruisme et la moralité comme effet dérivé. Il semblerait toutefois que ces auteurs conçoivent l'altruisme non en termes de motivation mais en termes de motifs. Or il faut distinguer la question de l'évolution de l'altruisme psychologique motivationnel de celle de l'altruisme psychologique sophistiqué ; nous verrons qu'il y a de bonnes raisons de croire que le second, mais non le premier dont il est question dans cette section, est un effet dérivé.

[200] D'autres auteurs défendent de manière plus ou moins explicite une théorie du produit dérivé mais font reposer l'altruisme psychologique sur d'autres capacités. Thomas NAGEL (1970) et Peter SINGER (1981) lient l'altruisme à la raison. Toutefois, outre la difficulté de proposer une explication évolutionnaire de ce concept flou qu'est la raison, il semblerait que ces auteurs conçoivent l'altruisme psychologique dans sa conception sophistiquée. Or c'est l'altruisme motivationnel qui nous intéresse ici. Je ne m'attarderai donc pas sur ces positions.

[201] Ce qui est moins clair est de savoir si elle a été sélectionnée pour permettre de mieux tricher ou de mieux coopérer (à ce propos, voir plus haut, note 167 ; H. MOLL et TOMASELLO 2007).

[202] L'exercice de l'empathie correspond plus ou moins à « se mettre à la place d'autrui » (D. KREBS & RUSSELL 1981). Cette capacité cognitive est à la base des émotions empathiques comme la compassion, la pitié ou la sympathie.

phénomènes seraient liés parce qu'ils découlent tous deux de l'activation de neurones miroirs. Ces derniers sont une sorte particulière de neurones qui deviennent actifs à la fois lorsque l'on effectue nous-mêmes une action et lorsque l'on observe quelqu'un d'autre produire une action similaire[203] (ou lorsque l'on ressent nous-mêmes une émotion et lorsque l'on observe les manifestations de cette émotion chez quelqu'un d'autre). Les neurones miroirs permettent en quelque sorte de se mettre en phase avec autrui. Selon un certain nombre d'auteurs, des phénomènes comme l'imitation et la compréhension des actions (CRAIGHERO & RIZZOLATTI 2004), l'empathie (PRESTON & DE WAAL 2002/2001 ; GALLESE *et al.* 2004) et la théorie de l'esprit (GALLESE & GOLDMAN 1998) reposent sur l'activation de ces neurones miroirs. Ces derniers auraient été sélectionnés précisément parce qu'ils remplissent ces diverses fonctions. Au fond il se pourrait bien que l'empathie cognitive corresponde à la théorie de l'esprit dans son mode émotion, ou appliquée aux émotions.

D'autre part, l'émotion de sympathie repose directement sur le mécanisme de l'empathie si bien que les deux sont très souvent corrélés: lorsque nous voyons quelqu'un souffrir, nous comprenons son état émotionnel et souffrons aussi (évidemment dans une moindre mesure);[204] de plus, ce sentiment négatif nous incite à entreprendre quelque chose pour atténuer la souffrance de cette personne.

En bref, la sympathie et la théorie de l'esprit seraient liées par les neurones miroirs, via l'empathie. Les neurones miroirs auraient été sélectionnés parce qu'ils apportent des effets bénéfiques au niveau des choix stratégiques via la théorie de l'esprit; mais lire dans l'esprit des gens signifie aussi saisir et ressentir dans une certaine mesure leurs émotions. La sympathie serait donc un effet dérivé qui n'apporte aucun avantage sélectif, bien au contraire.

Il me semble que cette théorie du produit dérivé est difficile à défendre. D'abord, la thèse selon laquelle la théorie de l'esprit reposerait directement sur les neurones miroirs est controversée. Comme le notent Pierre JACOB et Marc JEANNEROD (2004), les neurones miroirs relèvent d'un phénomène sensori-moteur qui permet à la rigueur de comprendre une action observée mais ne semble pas fournir les clés de la

[203] Les expériences originales ont été faites sur des macaques (GALLESE *et al.* 1996).

[204] Dans la réalité, l'empathie et la sympathie sont des phénomènes difficiles à distinguer si bien qu'ils sont souvent associés voire même confondus dans la littérature. Ainsi, le mot « empathie » est souvent utilisé pour désigner à la fois la capacité de comprendre les états affectifs d'autrui, et celle de partager, ou plutôt se mettre en phase avec les émotions et sensations d'autrui (ce que j'appelle sympathie).

compréhension du contenu de la pensée des gens ; la théorie de l'esprit ne reposerait donc pas directement sur les neurones miroirs. D'autre part, étant donné que nous disposons déjà d'une série d'explications congruentes (et à mon avis convaincantes) en termes d'avantages sélectifs de l'évolution de l'altruisme psychologique, je ne vois pas vraiment de raison d'élaborer une théorie du produit dérivé. Enfin, cette théorie réduit l'altruisme psychologique à l'émotion de sympathie. Or nous avons vu qu'il existe d'autres candidats valables comme l'amour ou l'amitié.

En revanche il se pourrait bien que les neurones miroirs soient à la base de l'émotion de la sympathie; nous disposerions ainsi d'un complément d'information (entièrement compatible avec les interprétations proposées précédemment) sur l'évolution de cette émotion altruiste particulière.

Conclusion

Ce chapitre avait pour but de montrer, contre les défenseurs de la thèse de l'égoïsme psychologique, que l'altruisme psychologique n'est pas une chimère. Ma démonstration repose sur une simplification du débat. J'ai commencé par constater que le débat autour de l'altruisme psychologique a été envahi par un vocabulaire et une réflexion en termes de désirs, buts ou intentions. A mon avis, cette perspective voile le problème et fait oublier qu'au fond, la question essentielle de la controverse est celle de la motivation. Sur cette constatation, j'ai établi une distinction entre une forme sophistiquée et une forme motivationnelle de l'altruisme psychologique, suggérant de focaliser l'attention sur la seconde. Au fond l'altruisme motivationnel signifie simplement que l'on peut être touché par ce qui concerne le bien-être et les intérêts d'autres personnes, au point d'être motivé à agir en leur faveur. Vu sous cet angle, la thèse de l'égoïsme psychologique n'est plus tenable. Les données empiriques issues de la psychologie et les explications de l'évolution de l'altruisme psychologique achèvent de montrer que nous sommes bel et bien capables de produire des actions altruistes psychologiques.

Le second objectif majeur de ce chapitre était de définir les dénominateurs communs entre l'altruisme évolutionnaire et l'altruisme psychologique. Au-delà du simple fait que les deux formes d'altruisme concernent la promotion du bien-être et des intérêts d'autrui au détriment du bien-être et des intérêts propres de l'agent, au vu des

explications évolutionnaires qui ont été présentées, il ressort clairement que le lien entre les deux formes d'altruisme est à trouver dans l'explication de leur genèse. Très schématiquement, la plupart des explications évolutionnaires proposées dans ce chapitre procèdent de la manière suivante. Les théories évolutionnistes permettent d'expliquer pourquoi des types de comportements altruistes évolutionnaires sont apparus, se sont répandus et ont été maintenus au fil de l'évolution humaine. Ces mêmes théories permettent de saisir l'impact des comportements altruistes évolutionnaires sur l'environnement social des êtres humains. Il se trouve que cet impact est propice à l'apparition et à la diffusion de tendances psychologiques à l'altruisme psychologique. Ainsi, il est envisageable d'établir un lien historique et évolutionnaire entre les deux formes d'altruisme. Plus précisément, la sélection et la propagation de l'altruisme évolutionnaire est une condition nécessaire à l'évolution de l'altruisme psychologique. Si ce dernier est apparu et s'est maintenu au cours de l'histoire humaine, c'est parce qu'il a pour fonction de soutenir l'altruisme évolutionnaire.

En définitive, la raison ultime des actions altruistes (aux deux sens du terme) est le maintien de la coopération, laquelle favorise de manière plus ou moins directe le bien-être des individus qui la pratiquent.

Il reste encore à élucider le rapport entre l'altruisme et la morale. Dans le chapitre suivant, il apparaîtra que la forme motivationnelle de l'altruisme psychologique est essentielle pour nous motiver à l'action morale. De plus, en tant que réaction émotionnelle face à certains types de situations, elle fournit les premières impulsions à la réflexion morale proprement dite. Quant à l'altruisme sophistiqué, je suggèrerai qu'il s'agit d'une condition nécessaire à la production d'assertions morales.

Seconde partie

4. Ethique, morale et ambitions de l'éthique évolutionniste

Après cette première partie qui relève plutôt de la philosophie des sciences, les questions propres à la philosophie morale peuvent être abordées. L'approche qui sera développée dans la deuxième partie de cet ouvrage sera de nature empirique et évolutionnaire et s'inspirera largement des théories vues jusqu'ici. Ce court chapitre a pour objectif de poser quelques définitions fondamentales et une structure sur laquelle reposent les chapitres suivants.

Les termes d'« éthique » et de « morale » sont définis d'innombrables façons. C'est la raison pour laquelle il me paraît nécessaire d'éclairer le lecteur sur le sens que je leur donne. Dans ce qui suit, je commencerai par présenter ce qui me semble être les quatre grands domaines de l'éthique avant de proposer une première approximation de la moralité. Ces définitions et catégorisations me permettront de donner un premier aperçu de l'apport potentiel en éthique des théories évolutionnistes et des données expérimentales. Les bases seront ainsi posées pour entamer l'élaboration détaillée d'une éthique évolutionniste.

4.1. Les quatre niveaux de l'éthique

Par « éthique », j'entends le domaine de recherche qui concerne toutes les questions en rapport avec la pensée et les comportements moraux. Globalement, l'éthique couvre quatre grands niveaux de réflexion : l'éthique descriptive, la métaéthique, l'éthique normative et l'éthique appliquée. Voyons en détail de quoi il s'agit :

Au niveau de l'éthique descriptive, on s'intéresse à l'acquisition, au maintien et au fonctionnement des capacités humaines nécessaires à la moralité ainsi qu'à la dynamique de la morale au sein des sociétés : Comment la pensée et les comportements normatifs (et plus particulièrement moraux) sont-ils apparus au cours de l'histoire ? Quelle est la fonction évolutionnaire de la moralité (s'il y en a une) ? Quel est le rôle de la morale dans une société ? Quels sont les normes ou valeurs généralement admises

dans les sociétés humaines ? Peut-on parler de sens moral ? L'éthique descriptive s'intéresse également aux questions d'ordre psychologique : Qu'est-ce qu'une expérience morale ? Quels pourraient-être les facteurs qui influencent la pensée et les comportements moraux ? Quels sont les rôles respectifs tenus par les émotions et la raison dans la pensée et l'activité morales ? Qu'est-ce qui nous motive à agir moralement ? Comment la moralité se développe-t-elle au cours de l'ontogenèse des individus ?

Le second domaine est la métaéthique. Je me réfère ici à la tradition de philosophie morale analytique qui s'est donnée pour objet de trouver des réponses à des questions de nature prénormative. De manière assez schématique, la métaéthique peut être divisée en trois types de questions. i) Il y a les questions d'ordre ontologique. Quelle est la nature du rapport entre ce qui est moral et ce qui ne l'est pas ? Qu'est-ce qu'une propriété morale ? Les propriétés morales existent-elles indépendamment de nos croyances et de nos attitudes ? Si ce n'est pas le cas, sont-elles le pur produit de notre esprit ? ii) Il y a les questions d'ordre sémantique. Quelle est la signification des énoncés moraux ? Les assertions morales (normes, jugements) sont-elles susceptibles de vérité ? iii) Enfin, il y a les questions d'ordre épistémique. Peut-on parler de connaissance morale ? Si c'est le cas, quelle est la nature de cette connaissance morale et comment est-elle acquise ?

La limite entre le domaine descriptif et le domaine métaéthique est un peu arbitraire. En tous les cas elle ne peut pas être tracée de manière très nette. Certaines questions sont traitées aux deux niveaux. Par exemple, celle de savoir s'il existe un sens moral inné ou si la moralité est plutôt le fruit de l'apprentissage. On pourrait penser que l'éthique descriptive est en réalité une partie de la métaéthique. Stélios VIRVIDAKIS (1996, p. 13-14) par exemple, classe les questions d'ordre psychologique dans le domaine de la métaéthique. Michael BRADIE (1994) prend la même position sur les questions de l'origine, de la fonction et du développement de la moralité.[205] Il me paraît toutefois plus éclairant de séparer ces deux domaines de recherche pour la simple raison que le premier ressortit en grande partie de la recherche scientifique ou des « sciences du terrain » ; il nécessite un recours constant aux théories et données évolutionnaires,

[205] « I employ the term meta-ethics in a broad sense to include not only these traditional questions of justification, objectivity, and meaning, but also questions about the origin, function, and development of morality. » (BRADIE 1994, p. 12)

aux sciences cognitives en général, ainsi qu'à l'anthropologie et à la sociologie. La métaéthique en revanche relève plutôt de la tradition de l'argumentation philosophique.

Le troisième domaine important de la réflexion éthique est celui de l'éthique normative. On y développe les systèmes moraux : plus précisément, on définit ce qui est bien et ce qui est mal, on fonde les valeurs et principes fondamentaux d'une théorie morale et on justifie les normes et jugements moraux. L'éthique normative est aussi l'étude systématique et comparative de conceptions morales ou de systèmes moraux comme l'utilitarisme, le kantisme, les théories des vertus, etc.

Certains auteurs considèrent la question de la justification des assertions morales comme relevant également de la métaéthique (RUSE 1993 p. 53 ; BRADIE 1994, p. 12 ; RAUSCHER 1997, p. 305). Je considère cela comme une ingérence malvenue dans le domaine de l'éthique normative car ce faisant, cette dernière se retrouverait vidée de toute substance. Il est clair que certaines prises de position au niveau métaéthique ont des implications directes au niveau normatif, mais cela ne me paraît pas être une bonne raison pour retirer à l'éthique normative la problématique de la justification des assertions morales.

Le quatrième et dernier niveau de la réflexion éthique est celui de l'éthique appliquée où l'on cherche à résoudre les conflits moraux existants (par exemple les controverses autour de l'euthanasie, du clonage, du droit des animaux, etc.) notamment en appliquant des normes et principes préalablement acceptés.[206]

Il convient de remarquer que malgré la commodité de cette classification, les frontières entre ces différents niveaux ne sont pas nettes et qu'une prise de position à un niveau peut largement influencer le développement des autres.

Dans la suite de cet ouvrage un chapitre sera consacré à chacun des trois premiers niveaux de réflexion, laissant de côté celui de l'éthique appliquée.[207] Mais avant cela, un certain nombre de définitions et mises en place des problématiques s'impose.

[206] Notons qu'il existe également des courants d'éthique appliquée qui ne cherchent pas forcément à s'appuyer sur des théories morales existantes pour régler les dilemmes moraux mais suivent de préférence une procédure casuistique par comparaison avec des cas analogues paradigmatiques sur lesquels il y a consensus. A ce propos, voir JONSEN & TOULMIN 1988.

[207] Ne pas traiter d'éthique appliquée dans le cadre de ce travail est un choix motivé par la volonté de délimiter quelque peu le champ de ma recherche.

4.2. Une première approximation de la moralité

Dans la section précédente, j'ai soutenu que l'éthique est le domaine de recherche qui concerne toutes les questions en rapport avec la pensée et les comportements moraux. Il s'agit donc d'expliquer ce que j'entends par « moral » ou « moralité ».

4.2.1. Une définition sommaire

Il y a à peu près autant d'interprétations des termes « moralité » ou « moral » que de théories défendues en éthique ; leur définition précise dépend directement de prises de position aux quatre niveaux de l'éthique. Ainsi, ces notions ne pourront être précisées qu'au fil du travail. Pour le moment, commençons par une première approximation qui sera complétée au chapitre suivant (section 5.4).[208]

De manière très générale, la moralité concerne nos actions et les motifs qui nous poussent à accomplir ces actions. Elle est également liée à des valeurs et à des normes de conduite. Ces valeurs et normes sont assorties d'une valence objective et d'une obligation de les respecter. Plus précisément, la moralité est liée d'une part à une propension à considérer les valeurs et les normes auxquelles on adhère comme intersubjectivement valables, d'autre part à la notion de prescription, c'est-à-dire à une attente de comportement conforme aux valeurs et normes de conduite considérées comme morales et un désir de sanctionner en cas de transgression. Ces deux processus (adhésion subjective et attente de conformité) sont émotionnellement chargés. D'autre part, on attend des valeurs et des normes de conduite considérées comme morales qu'elles soient fondées et qu'elles puissent être acceptées par autrui. La moralité est donc liée à un acte réflexif ; les agents moraux doivent pouvoir délibérer, réfléchir sur les valeurs et normes qu'ils prônent et sur les actions qui doivent être faites. Toutefois, comme nous le verrons au chapitre suivant (section 5.2.1), cela n'empêche pas que cette réflexion puisse être largement guidée par nos émotions et par des biais psychologiques.

Enfin la moralité est liée au rapport du sujet avec autrui (plus précisément, je défendrai dans la section suivante et 5.4.1 que ce rapport est de type altruiste) ainsi qu'à des notions caractéristiques telles que le « bien » ou le « mal », lesquelles évidemment,

[208] De plus, il est difficile de donner une définition de la moralité qui ne fasse pas appel, de manière plus ou moins évidente, à la notion même de moralité. Je tâcherai ici d'éviter cette circularité.

trouvent des significations fort diverses selon les contextes culturels (et pour les philosophes, selon les théories éthiques).

Cette définition approximative ne convaincra pas tout le monde. Par exemple, certains auteurs pensent que la moralité est une simple affaire de sentiments et qu'il n'est pas nécessaire de postuler l'activité réflexive des sujets moraux (WALLER 1997). D'autres affirment qu'un aspect fondamental de la moralité est que toutes les normes qu'elle prescrit possèdent une prétention, non seulement à une forme d'objectivité mais à l'universalité (KANT 1997/1785 ; JACKSON 1998 ; RUSE 1998, p. 69[209]). Toutefois, je ne suis pas certaine que ces deux voies correspondent réellement à la manière dont les gens conçoivent habituellement la morale (et les « gens » ne se restreignent pas aux philosophes). Faire de la morale une simple affaire de sentiments revient à attribuer à certains animaux le statut d'être moral. Il faudrait alors blâmer moralement le chimpanzé chapardeur… Quant à l'idéal universaliste, il semble davantage relever de la construction de philosophique. C'est du moins ce que laissent entendre les résultats d'un certain nombre de tests empiriques ; ces études attestent en effet que beaucoup de gens ne sont pas prêts à considérer que leurs normes morales ont une prétention à l'universalité (à ce propos, voir NICHOLS 2004, pp. 669-171, STICH & WEINBERG 2001, RYAN 1997 ; voir aussi les expériences de KELLY et HUSSARD dont je rends compte à la page 298). La thèse plus modeste d'une propension (plus ou moins forte selon les cas) à prétendre à la valeur intersubjective de nos convictions morales me paraît moins problématique.

En définitive, il me semble qu'en refusant l'une ou l'autre composante de ma définition de la moralité, nous nous éloignerions trop de la conception ordinaire que les gens se font de ce terme.

4.2.2. Le rapport entre la moralité et l'altruisme

Un certain nombre d'auteurs défend l'idée que la moralité peut être réduite à l'altruisme. Voyons dans quelle mesure cette position peut être prise au sérieux.

Nous avons vu qu'il existe deux sortes d'altruisme logiquement distinctes l'une de l'autre (section 3.2) : la version évolutionnaire et la version psychologique. Parler

[209] Précisons qu'à la différence de KANT ou JACKSON, RUSE pense que le caractère universel de la morale ne prend sens que dans l'esprit des gens.

d'altruisme évolutionnaire est pertinent uniquement lorsque l'on considère des types de comportements et non des actions particulières. D'autre part, cette forme d'altruisme est relative aux effets d'un comportement mais ne tient pas compte des motivations des agents. Or la moralité porte sur des actions particulières et sur les motifs qui ont poussés les agents à accomplir leurs actions. Il paraît donc évident que l'altruisme évolutionnaire n'est pas directement lié à la moralité. Nous avons également vu (section 3.3.3) que l'altruisme psychologique se décline sous deux formes selon que l'on considère la question de la motivation à l'action (*altruisme psychologique motivationnel*) ou ce que les gens conçoivent comme étant les motifs de leurs actions (*altruisme psychologique sophistiqué*). Ainsi, il semblerait que la moralité peut être mise prioritairement en rapport avec l'altruisme psychologique sophistiqué puisque ce dernier concerne précisément les motifs des actions particulières.[210]

Certains tenants de l'éthique évolutionniste établissent une équivalence entre une action morale et une action altruiste, cette dernière étant comprise au sens d'une action motivée par la volonté de coopérer et de faire du bien à autrui. Robert RICHARDS (1986) par exemple assimile la moralité et l'intention altruiste qui, selon lui, consiste dans le fait de chercher à promouvoir le bien-être de la communauté et de ses membres.[211] De manière similaire, Michael RUSE (1998) pense que nous possédons un sens moral qui s'exprime à travers des sentiments altruistes,[212] ces derniers devant être compris au sens de motivation à coopérer et à aider autrui. D'autre part, ces auteurs conçoivent un lien étroit entre l'altruisme évolutionnaire et l'altruisme psychologique : le motif altruiste psychologique jouerait le rôle d'une cause proximale qui mène à des actions altruistes évolutionnaires stables du point de vue évolutionnaire. En d'autres termes, pour agir de

[210] L'altruisme motivationnel est également lié à la moralité mais nous verrons au chapitre 5 qu'il ne peut pas en constituer un critère nécessaire. En revanche, il est essentiel pour nous motiver à agir moralement et il guide nos choix évaluatifs.

[211] « They [les êtres humains] have been formed 'to regard and advance the community good' and approve of altruism in others. (…) Having such a set of attitudes and acting on them is what we mean by being moral. » (RICHARDS 1986, p. 289) Notons que RICHARDS mélange les deux formes d'altruisme psychologique.

[212] Quoiqu'il ne s'exprime pas très clairement sur le sujet, RUSE semble le plus souvent penser en termes d'altruisme motivationnel plutôt que d'altruisme sophistiqué.

manière altruiste au sens évolutionnaire, nous devons être altruistes au sens psychologique.[213]

Quel crédit peut-on accorder à cette idée de réduction de la morale à l'altruisme ? Au-delà du fait que ce genre de positions confond souvent la forme motivationnelle et la forme sophistiquée de l'altruisme (section 3.3.3), un certain nombre de critiques peuvent être adressées contre cette idée :

Tout d'abord, les partisans de l'égoïsme psychologique (HOBBES 2000/1651 ; CIALDINI *et al.* 1987) rejetteraient en bloc l'idée d'assimiler la morale à l'altruisme pour la simple raison que les êtres humains sont incapables de former des motifs réellement altruistes. Mais comme nous l'avons vu au chapitre 3 (section 3.3), si l'on considère la question de la motivation (qui est pertinente dans le contexte de la controverse entre les défenseurs de l'égoïsme psychologique et leurs opposants), la position égoïste psychologique n'est pas tenable. Je ne m'y attarderai donc pas.

Une autre manière de rejeter l'identification des actions morales aux actions altruistes est de défendre une position conséquentialiste, selon laquelle peu importe les motifs sous-jacents, seules les conséquences importent pour déterminer la valeur morale des actions. Ainsi l'altruisme ne serait pas un ingrédient nécessaire à la moralité. La définition de la morale proposée plus haut ne permet pas de contrer cette approche mais nous verrons que les théories conséquentialistes ne font pas forcément bon ménage avec la conception particulière de la moralité élaborée plus loin à la section section 5.6. Pour cette raison, je ne m'étendrai pas ici sur cette critique.

Une autre objection contre cette identification est de chercher à montrer que les domaines de l'altruisme et de la moralité se recoupent sans se recouvrir. Elliott SOBER (1993) par exemple affirme d'une part que toute action altruiste n'est pas forcément morale et d'autre part que presque tous les systèmes moraux requièrent que les individus agissent parfois dans leur propre intérêt plutôt qu'en faveur de celui d'autrui (ROTTSCHAEFER et MARTINSEN défendent également cette idée : 1990, pp. 152-153).[214] Pour soutenir son propos, SOBER propose le cas de figure suivant. Imaginez que vous possédez un médicament rare. Vous êtes très malade et la prise de ce médicament est

[213] « We are moved by genuine, non-metaphorical altruism. To get 'altruism' [au sens évolutionnaire], we humans are altruistic [au sens psychologique]. » (RUSE 1998, p. 222; voir aussi RUSE 1986, p. 104)

[214] « Evolutionists have emphasized that part of morality that requires us to behave altruistically. But no morality requires limitless altruism. Almost all require that an individual sometimes places self ahead of other. » (SOBER 1993, p. 213)

une question de vie ou de mort. D'autre part, une autre personne à vos côtés serait également susceptible de profiter de ce médicament à la différence que sa vie n'en dépend pas (elle pourrait très bien se soigner par d'autres moyens quoique de manière moins efficace). Supposons que le médicament ne peut pas être divisé entre vous deux ; il ne peut être ingéré que par une personne. Dans ces conditions, SOBER se demande quelle est l'action moralement correcte. Selon lui, la morale dicterait que vous êtes autorisé à prendre le médicament pour vous. Cela signifie que la morale requiert que vous fassiez passer votre intérêt individuel avant celui d'autrui. En quelque sorte, vous agiriez à la fois de manière égoïste et moralement bonne. SOBER en conclut que la moralité n'est pas quelque chose qui s'oppose à l'égoïsme.[215] Toutefois, cet argument ne me paraît pas pertinent car il pose une mauvaise question. Dans ce cas de figure, il ne s'agit pas de savoir ce qu'il est moralement correct de faire, mais ce qu'il est permis de faire. En l'occurrence, nous pouvons dire qu'il est permis d'agir dans notre intérêt personnel sans pour autant contrevenir à nos convictions morales ; en nous arrogant le médicament, nous ne causons pas de réel tort à notre voisin, lequel ne serait d'ailleurs pas justifié à se sentir lésé dans ces circonstances. En conséquence, l'exemple du médicament ne peut pas être utilisé en faveur de la thèse que toute action morale n'est pas forcément altruiste. De manière générale, je doute que les défenseurs d'une position à la SOBER puissent fournir des exemples convaincants où la moralité *requiert* que l'on place ses intérêts propres et son bien-être avant ceux d'autrui. Même si elle *permet* certaines actions motivées par la recherche de ses intérêts et de son propre bien-être, la moralité ne semble en *prescrire* aucune. C'est du moins la conclusion à laquelle nous pousse la définition de la moralité proposée plus haut ; les valeurs et normes sont assorties d'une valence objective et d'une obligation à les respecter. Le domaine de la morale se situe donc au-delà du simple permissible.

SOBER considère d'autres situations qui selon lui sont altruistes mais non morales (SOBER & D. WILSON 2003/1998, p. 239). Il s'agit des cas où l'on applique des principes moraux sans tenir compte des résultats de l'application de ces principes sur les intérêts et le bien-être d'autrui. Par exemple, on pourrait imaginer que certaines personnes suivent à la lettre des principes simplement parce qu'ils ont été dictés par Dieu. A nouveau cet exemple me paraît peu convainquant. Il me semble qu'en suivant des principes édictés par Dieu, nous nous trouvons dans le domaine religieux et non

[215] « Morality is not something that stands in opposition to selfishness. » (SOBER 1993, p. 213)

moral. Ce point deviendra plus clair à la lumière de la section 5.4.3 où je montrerai que si l'on adopte mon analyse de la notion de norme, il faut admettre que ces personnes suivent en réalité des normes d'autorité et non des normes morales.

Dans la ligne de SOBER, un autre cas potentiel d'action morale non altruiste pourrait être le suivant. Lors d'un procès, un déontologiste kantien décide de témoigner en faveur de son voisin qu'il exècre au plus haut point. Il le fait pour la seule et unique raison qu'il souscrit à la maxime selon laquelle il ne faut jamais mentir ; en cela il obéit à l'impératif catégorique. Ce n'est pas par désir de soutenir son voisin (ce qui serait un exemple d'altruisme) mais par désir de suivre l'impératif catégorique que cet homme agit. Au-delà de la question de savoir s'il est possible d'avoir un désir de suivre l'impératif catégorique suffisamment fort pour causer l'action (les arguments que j'avancerai à la section 5.2.2, laissent plutôt penser que ce n'est pas le cas), je pense que ce genre d'exemples peut très bien être interprété en termes altruistes (au sens d'altruisme sophistiqué). En effet une formulation de l'impératif catégorique est que l'on considère autrui comme une fin et non comme un moyen[216]. Ne s'agit-il pas d'une manière de tenir compte des intérêts et du bien-être d'autrui ? Suivre des principes moraux sans tenir compte des résultats de l'application de ces principes ne signifie pas forcément que l'on s'éloigne de l'altruisme ; si les principes moraux sont choisis en fonction de considérations sur les intérêts et le bien-être d'autrui (c'est en ce sens que j'argumenterai à la section 5.4), alors on reste dans le domaine de l'altruisme.

RICHARDS et RUSE semblent donc toucher juste lorsqu'ils affirment que toute action morale est une action altruiste. Par contre, il n'est pas aussi évident que la converse soit le cas. Il semble possible que des actions liées à des motifs altruistes puissent être considérées comme moralement inacceptables (ce qui donnerait partiellement raison à SOBER dans son refus d'assimiler les actions altruistes à des actions morales). Une raison en est que la moralité ne semble pas requérir l'altruisme illimité. Quoiqu'admirables, les actions que l'on appelle « surérogatoires » (celles qui sont considérées comme admirables sans pour autant que l'on se sente obligé de les réaliser soi-même ; par exemple donner tout ses biens à une association caritative) ne doivent pas forcément être considérées comme morales (FELDMAN 1986). Mais il n'est

[216] « Agis de telle sorte que tu traites l'humanité aussi bien dans ta personne que dans la personne de tout autre toujours en même temps comme une fin, et jamais simplement comme un moyen. » (KANT 1997, p. 105)

même pas nécessaire d'aller chercher si loin. Considérons l'exemple suivant. Durant la Deuxième Guerre mondiale, Ralf cache une femme juive dans sa cave. Il sait que des nazis suspicieux s'apprêtent à fouiller sa maison et que s'ils découvrent cette femme, lui-même et toute sa famille seront tués. Le choix de Ralf peut clairement être classé dans la catégorie des actions altruistes, à la fois au sens motivationnel (la compassion pour cette femme juive le pousse à faire ce choix) et sophistiqué (Ralf a l'intention de protéger cette femme). Par contre, la moralité de son action peut être mise à caution au moins du point de vue des membres de sa famille dont il « hypothèque la vie ».[217]

En conclusion, mon avis est que la moralité ne se réduit pas à l'altruisme mais ce dernier est une condition nécessaire à la moralité ; le domaine de la moralité est englobé dans celui de l'altruisme.

4.2.3. Une moralité prescriptive

La prescription me paraît être un concept central en éthique[218]. Ce n'est pas qu'elle provienne du simple fait d'évaluer. En effet, il serait absurde de prescrire tout ce que nous évaluons positivement (par exemple la beauté ou le talent) et interdire tout ce que nous évaluons négativement. En revanche, la prescription est liée aux normes, plus précisément aux normes de conduite. Comme nous l'avons vu à la section 2.3.5, les normes sociales (ou plus précisément la capacité à formuler des normes sociales) sont apparues au cours de l'évolution précisément parce qu'elles sont prescriptives ; dans la mesure où cette prescription est opérante (c'est-à-dire qu'il y a effectivement sanction des déviances), elles peuvent jouer leur rôle de garant de la coopération et de la coordination à l'intérieur de moyennes et grandes communautés (GINTIS 2003 ; BOWLES *et al.* 2003).

De plus, nous verrons à la section 5.4.4, que c'est essentiellement à travers les normes morales (qui sont une sorte de norme de conduite) que l'on peut individuer la moralité. Ainsi la prescription (ou l'obligation) est un concept central pour l'éthique puisqu'il est directement lié aux normes morales, par nature prescriptives.

---

[217] Dans la même veine, Elliott SOBER et David WILSON proposent l'exemple de Alan qui vole la carte de crédit de Betty pour sortir Carl d'un mauvais pas financier (2003/1998, p. 239).

[218] Au contraire de ce que pourraient penser d'autres philosophes comme Elisabeth ANSCOMBE (1958).

Une précision importante s'impose ici. Si dans les faits, les normes sont effectivement corrélées à une motivation à punir les déviances, ce serait une erreur d'en déduire que ce sont les normes elles-mêmes qui nous motivent à la punition.[219] A la section 5.2.2 je défendrai l'idée que les normes n'ont en réalité aucun pouvoir motivant. La motivation vient en amont des normes ; par exemple, si nous désirons que les gens coopèrent dans certaines circonstances, ce désir nous fera concevoir la norme *et* nous motivera à punir les individus opportunistes.[220]

4.3. Les ambitions de l'éthique évolutionniste

Les données scientifiques ont contribué de manière significative (de façon directe ou indirecte) à répondre aux questions traditionnellement traitées en philosophie. L'utilisation des sciences a par exemple énormément enrichi les débats sur le temps, l'espace, la causalité ou la perception. Dès lors, il est légitime de se demander dans quelle mesure la pensée et l'action morales peuvent être mieux comprises si l'on tient compte des données des sciences naturelles et sociales. Pour ce qui est de la morale, les disciplines les plus pertinentes sont d'une part, les sciences sociales traditionnelles (psychologie, sociologie, anthropologie), d'autre part, les théories évolutionnistes (biologie, anthropologie, psychologie, théorie des jeux) et certains résultats empiriques obtenus notamment en neurologie ou en économie expérimentale. Les théories évolutionnistes retiendront particulièrement mon attention. Il y a deux manières paradigmatiques d'en envisager la pertinence pour la réflexion morale.

La première part d'un constat négatif : malgré plus de deux millénaires de recherche philosophique, aucun système moral efficace n'a pu s'imposer (pas même dans la communauté des philosophes de la morale !). Suivant cette ligne de pensée, Andrzej ELZANOWSKI écrit :

[219] Beaucoup d'auteurs (en particulier les acteurs de la seconde génération de la théorie des jeux comme GÄCHTER & FALK 2002) stipulent sans plus d'argument que lorsqu'elles sont acquises par un individu, les normes induisent des comportements conformes à ces normes ainsi que la sanction des opportunistes.

[220] Ainsi on peut expliquer comment il est possible d'adhérer à une norme sans être motivé à agir conformément à elle et à punir les comportements déviants.

> « La philosophie a produit un nombre considérable d'écoles de pensée incompatibles entre elles qui traitent des valeurs humaines et s'est montrée particulièrement impuissante en matière de justification de jugements de valeur fondamentaux non dérivatifs. (…) De plus, l'absence d'une théorie consistante des valeurs humaines les rendent vulnérables à la démagogie politique et religieuse. (…) Ne comptant pour le moment qu'une poignée de tentatives de synthèse théorique, la science des valeurs humaines est encore à un stade embryonnaire ; malgré cela l'approche scientifique semble bien plus prometteuse que la pure philosophisation. » (ELZANOWSKI 1993, p. 259, ma traduction)[221]

Face à cet échec, on propose d'amener de l'eau fraîche au moulin en considérant les données scientifiques pertinentes pour une réflexion éthique. Il s'agit évidemment d'une manière un peu abrupte de présenter l'idée que le philosophe de la morale aurait beaucoup à gagner de prendre en considération les données des sciences.

La seconde manière de présenter les choses, moins polémique, est de dire que si on accepte la théorie de l'évolution darwinienne, on ne peut pas échapper à ses implications au niveau philosophique ; la théorie de l'évolution nous pousse à considérer le comportement moral comme un phénomène entièrement naturel et indépendant des lois divines ou d'autres phénomènes mystérieux ou non naturels. DARWIN (2000/1871) lui-même défendait ce point de vue puisqu'il pensait que le sens moral humain est une faculté apparue au fil de l'évolution.

A l'époque de DARWIN déjà, cette idée que la pensée et le comportement moral humain dépendent en bonne partie de notre histoire évolutionnaire a séduit beaucoup de penseurs, si bien que l'on peut véritablement parler de la naissance d'un courant de pensée : l'« éthique évolutionniste ».[222] De nos jours, ce courant compte un très grand

[221] « … philosophy has produced a number of incompatible schools of thought on human values and has proved essentially helpless on the issue of justification of basic, nonderivative value judgments. (…) In addition, the lack of a consistent theory of human values leaves them freely accessible to religious and political demagoguery. (…) Although, with only a handful of attempts at any sort of theoretical synthesis, the science of human values is still at the embryonic stage, the scientific approach seems to be much more promising than pure philosophizing.» (ELZANOWSKI 1993, p. 259)

[222] Depuis DARWIN, bien des penseurs se sont arrogés des idées de Darwin pour soutenir les positions morales les plus diverses et incompatibles (a ce propos, voir RUSE 2000). Cet ouvrage ne relevant

nombre d'adeptes. Au fil des années, cette approche évolutionnaire de la morale s'est enrichie de données provenant d'autres sciences que la biologie. Désormais, on s'inspire également des données de la théorie des jeux évolutionnaires, de la neurologie, de l'économie expérimentale, ainsi que de l'anthropologie et de la psychologie évolutionnistes. Au fond, l'éthique évolutionniste ne porte peut-être plus si bien son nom puisque son domaine d'inspiration s'est élargi au-delà des théories strictement évolutionnaires. Elle est le lieu par excellence de l'interdisciplinarité où des chercheurs provenant de toutes sortes de disciplines mettent en commun leurs idées et leurs connaissances.

L'objectif de la deuxième partie de ce livre sera d'analyser la manière dont les données issues des sciences et en particulier les théories évolutionnistes peuvent être intégrées dans une réflexion éthique. Il s'agira de sonder les limites et les possibilités de l'éthique évolutionniste. Cette dernière, comme on l'aura compris, doit être considérée comme une « méthodologie » plutôt qu'une théorie unitaire ; ce genre d'approche ne doit en aucun cas être considéré comme une alternative aux philosophies morales traditionnelles.[223] Au contraire, il faut considérer les données scientifiques comme un terreau fertile dans lequel le philosophe peut puiser de nouvelles questions et de nouvelles réponses. De cette entreprise, le résultat n'est pas à proprement parler *une* éthique évolutionniste mais un ensemble de théories, pas forcément compatibles entre elles, qui partagent une même méthode.

La délimitation des quatre grands domaines de l'éthique proposée ci-dessus donne un cadre pour réfléchir au genre d'usage que les philosophes peuvent avoir des théories évolutionnistes et des données expérimentales.[224] Voyons comment ces éléments scientifiques peuvent être utilisés à chacun des quatre niveaux.

---

toutefois pas de l'histoire des idées, je ne m'attarderai pas sur les théories qui composent l'histoire en partie bien malheureuse de l'éthique évolutionniste.

[223] Considérer l'éthique évolutionniste comme une concurrente à la morale traditionnelle est une idée tellement farfelue qu'elle n'a guère été soutenue, à l'exception peut-être d'E. WILSON qui (semblerait-il par simple goût du scandale) a suggéré de retirer temporairement la morale des mains des philosophes (1975, p. 562). Ainsi, le problème n'est pas de savoir, comme beaucoup l'ont cru, si les théoriciens évolutionnistes peuvent ou non retirer la morale des mains des philosophes ; le problème est de savoir dans quelle mesure les données des théories évolutionnistes sont pertinentes dans l'entreprise de l'élaboration des normes et principes moraux.

[224] Pour des analyses similaires, voir KITCHER 1994 ; BRADIE 1994 ; ROTTSCHAEFER 1998.

Au niveau de l'éthique appliquée, les données scientifiques nous permettent de prendre des décisions éthiques informées. Par exemple, si l'on décrète qu'il est interdit de pratiquer l'euthanasie sur des patients qui ne sont pas voués à une mort prochaine due à une maladie incurable, on peut recourir aux connaissances de la médecine pour définir si un patient donné se trouve ou non dans une telle situation. Relativement à ces questions, personne ne doute de l'intérêt de prendre en considération les données scientifiques, puisque cela ne remet aucunement en cause la philosophie morale traditionnelle.

Au niveau de l'éthique descriptive, il est évident que les données scientifiques et les théories évolutionnistes jouent un rôle important. Nous avons vu dans la première partie de cet ouvrage qu'elles nous apportent une meilleure compréhension de la manière dont fonctionnent les interactions sociales ; ce faisant, elles nous fournissent un excellent cadre conceptuel dans lequel il est possible de penser l'évolution et le fonctionnement de phénomènes intimement liés à la morale comme les émotions morales (voir sections 5.3.2 et 5.3.3) ou les normes sociales (section 2.3.5).

L'utilisation des données scientifiques et des théories évolutionnistes à ce niveau de l'éthique ne pose généralement pas de problème aux philosophes à condition qu'il s'agisse d'expliquer des capacités extrêmement générales comme la conscience d'autrui, l'empathie ou la faculté du raisonnement. Les « grincements de dents » commencent lorsque l'on cherche à expliquer de manière scientifique certaines activités morales comme l'utilisation de normes ou la valorisation de certains objets ou états de fait plutôt que d'autres. Au chapitre 5, il deviendra clair que non seulement il n'y a pas lieu de s'en horrifier, mais qu'au contraire, ce sont des investigations dans lesquelles tout philosophe de la morale devrait se lancer.

D'autre part, il est important de relever le fait que les données et théories élaborées au niveau de l'éthique descriptive peuvent exercer une influence non négligeable dans les autres domaines de l'éthique. Pour ce qui est du niveau de l'éthique normative, nous avons vu (section 3.3.4) que les êtres humains sont capables de motivations altruistes, c'est-à-dire que nous sommes dotés d'une tendance psychologique à aider non seulement nos proches parents mais également des personnes inconnues sans attente de service en retour, c'est-à-dire sans motivation égoïste sous-jacente. Sachant cela, toute théorie morale reposant sur une conception de l'être humain comme un calculateur égoïste devient douteuse : par exemple HOBBES (2000/1651),

voire même l'ensemble de la tradition morale contractualiste représentée par ROUSSEAU (2001/1762) ou RAWLS (1997/1971).

Pour ce qui est du niveau métaéthique, si les recherches en éthique descriptive permettent de montrer de manière convaincante que la morale est un produit de l'évolution,[225] sachant que l'évolution ne se dirige pas vers un but particulier (tous les biologistes reconnus s'accordent sur le fait que l'évolution n'a pas de dessein), il s'ensuit que les valeurs et les normes prônées par les hommes auraient pu être tout autres si l'évolution avait pris une autre direction. En conséquence, il y a tout lieu de mettre en doute les positions métaéthiques selon lesquelles les propriétés morales existeraient dans le monde de manière fixe et indépendante des croyances et attitudes des gens. J'argumenterai en ce sens au chapitre 6.

Etonnamment on trouve parmi les philosophes (par exemple KITCHER 1994[226]) une certaine réticence à admettre que les données scientifiques puissent nous donner des clés pour saisir la nature des valeurs et des énoncés moraux et aborder la question de la possibilité d'une connaissance morale. Au chapitre 6, il deviendra pourtant évident à quel point l'approche évolutionnaire peut porter ses fruits au niveau métaéthique.

Pour revenir à l'éthique normative, les théories évolutionnistes et les données empiriques peuvent s'avérer utiles de plusieurs manières. Elles peuvent par exemple, en conjonction avec certains principes moraux que l'on accepte par ailleurs, *inspirer* l'élaboration de nouvelles normes morales. Par exemple, s'il peut être montré que les jeunes de moins de vingt ans possèdent une forte tendance (extrêmement difficile à maîtriser) à prendre des risques immodérés, cela montrerait qu'ils ne peuvent pas être responsables de leurs actions au même titre que les personnes d'âge mûr. Sur la base de ces considérations (et si l'on accorde par ailleurs une valeur morale à la vie) on peut en venir à édicter une norme qui interdit aux moins de vingt ans de pratiquer des activités dans lesquelles la prise de risques inconsidérés peut mettre autrui en danger (par

[225] Au chapitre 5 (sections 5.1 et 5.5), je tâcherai de montrer que même si elle n'est pas une adaptation biologique, la moralité est néanmoins un produit (dérivé) de l'évolution.

[226] Il est à noter que sur ce point, KITCHER a récemment révisé sa position : « Although I once believed that the project of uncovering the historical processes through which modern moral practices have emerged, tracing them to distant biological roots, was perfectly legitimate, I didn't view it as affecting either philosophical discussions in meta-ethics or substantive normative debates. I now think that that view was mistaken on the first score: understanding the genealogy of morals has meta-ethical implications. » (1998, p. 306)

exemple la conduite de véhicules relativement puissants, la conduite de parapentes en tandem, etc.).

On peut également recourir aux données empiriques et théories évolutionnistes pour *préciser* le contenu de certains critères de justification morale. Par exemple, si on décide d'adopter un critère de justification morale qui intègre la notion de nature humaine (FOOT 2002), on peut recourir aux théories évolutionnistes pour découvrir les caractéristiques propres à cette nature humaine (les hommes sont des êtres sociaux qui dépendent les uns des autres pour leur survie, sont enclins à la colère ou à la compassion dans telles ou telles circonstances, ont besoin de telles ou telles conditions pour développer des comportements d'entraide, etc.).

Inversement, les données empiriques peuvent aider à *sélectionner* les systèmes moraux viables. Par exemple, si l'on dispose de connaissances précises relatives aux limites de certaines capacités humaines, il est possible de poser des restrictions aux obligations morales imposées aux êtres humains (à condition bien sûr d'admette que tout système moral crédible se doit d'être praticable). Imaginons que des données psychologiques permettent de montrer l'effet dévastateur du sentiment de responsabilité illimitée sur la santé psychique des gens ; dans ce cas, une morale prônant ce principe (JONAS 2001/1990) serait à rejeter. Souvenons-nous également de la possibilité mentionnée ci-dessus, de rejeter les systèmes moraux basés sur une conception psychologique égoïste de l'être humain.

Un autre apport extrêmement intéressant des théories et données scientifiques au niveau de l'éthique normative concerne la *justification* des normes et principes moraux. Cette idée est régulièrement fustigée dans la littérature (KITCHER 1994; NAGEL 1983; P. WILLIAMS 1993, ALEXANDER 1987 ; HUNEMAN 2007) ; je tâcherai cependant de montrer au chapitre 7 qu'il vaut la peine de la prendre au sérieux.

Conclusion

La grille de lecture et les ambitions de l'éthique évolutionniste étant posés, nous pouvons entrer dans le vif du sujet. Les chapitres suivants seront consacrés à trois domaines de l'éthique : l'éthique descriptive, la métaéthique et l'éthique normative. J'y développerai une conception personnelle de l'éthique évolutionniste qui doit beaucoup aux travaux antérieurs réalisés dans cette discipline. Par souci (au fond très arbitraire) de

délimiter quelque peu le sujet de ma recherche, l'éthique appliquée sera malheureusement mise au banc des laissés pour compte. Cela dit, il est clair que les prises de position aux trois autres niveaux de l'éthique sont susceptibles d'exercer un impact considérable sur le traitement des problèmes d'éthique appliquée. Voilà un sujet passionnant qui mériterait que l'on y consacre un ouvrage entier…

5. Ethique descriptive

Dans ce chapitre, je commencerai par passer en revue la manière dont les tenants de l'éthique évolutionniste ont abordé la question de l'éthique descriptive, c'est-à-dire leurs spéculations sur la naissance et le fonctionnement de la moralité. Nous verrons que les tentatives d'expliquer l'évolution de la morale conçue comme objet de sélection sont assez peu convaincantes. Une voie plus prometteuse est de considérer la morale comme un produit dérivé qui « surferait » sur un certain nombre de capacités et de biais psychologiques[227] qui ont évolué pour des raisons propres. Mais pour savoir sur quels biais et capacités repose la morale, il faut disposer au préalable d'une conception précise de l'activité morale. Nous verrons à quel point cette tâche est ardue ; elle ne pourra être réalisée qu'en plusieurs étapes. La première consistera à présenter une description précise et empiriquement informée de l'activité évaluative et normative ; c'est ce que j'appellerai le « tableau affectif ». Etant donné que l'activité morale est un cas particulier de l'activité évaluative et normative, le tableau affectif ne permettra pas de préciser tous les éléments cognitifs et psychologiques sous-jacents à la moralité. Il s'agira donc de persévérer dans la recherche des critères d'individuation (conditions nécessaires et suffisantes) de la moralité. Une tentative reposant sur la notion d'émotion morale s'avèrera peu concluante. En fin de compte, le grain spécifique de l'activité morale sera défini en fonction de la manière dont nous justifions nos jugements évaluatifs ; deux critères de la moralité permettront d'établir la liste des capacités et biais psychologiques sur lesquels elle repose.

5.1. Spéculations sur la genèse de la moralité

Les recherches anthropologiques montrent que toutes les sociétés humaines possèdent la moralité (S. ROBERTS 1979 ; BROWN 1991). Cette réalité nous incite à concevoir ce phénomène comme un produit de l'évolution qui peut être expliqué en termes d'adaptation et d'avantages sélectifs. D'un point de vue évolutionnaire, la moralité se traduit généralement en termes d'investissements produits par les individus

[227] Comme nous l'avons vu à la section 1.2.3 (p. 43), la notion de biais psychologique est utilisée ici au sens de tendance psychologique sans qu'aucune connotation négative n'y soit associée.

moraux en faveur du bien-être et des intérêts d'autres individus ou de la communauté entière. Ainsi, la moralité est à première vue coûteuse au niveau individuel. Puisqu'on part du principe qu'elle a évolué, deux voies théoriques s'ouvrent à nous. La première consiste à dire que la moralité n'est pas neutre du point de vue de la sélection, c'est-à-dire qu'elle apporte un avantage sélectif (au niveau génétique, individuel ou du groupe). La seconde voie considère la moralité comme un effet dérivé d'une ou plusieurs autres adaptations ; au contraire de ces dernières, la moralité n'aurait pas été sélectionnée pour un avantage sélectif dont elle serait la cause.

Si l'on opte pour la seconde solution, on pensera que la moralité est un produit dérivé qui repose sur des capacités qui ont évolué de manière propre (ROTTSCHAEFER & MARTINSEN 1990 ; ROTTSCHAEFER 1998/1997 ; SINGER 1981 ; PRINZ 2008) ; par exemple la capacité de réfléchir, de comprendre les états d'esprits d'autrui, de se mettre à la place d'autrui, de communiquer par le biais du langage, ou plus précisément les mécanismes liés à la sélection de parentèle ou à la réciprocité. Ainsi George WILLIAMS écrit :

> « Les systèmes éthiques (…) doivent avoir été produits de manière indirecte, par une sorte d'accident, chose qui se passe couramment dans l'évolution. (…) Ces motivations [morales] doivent provenir d'attitudes biologiques normales favorisées par la sélection de parentèle ou la réciprocité, mais ont des manifestations anormales dans le cadre de notre environnement moderne anormal. C'est de cette anormalité que dépend l'éthique humaine. » (G. WILLIAMS 1993, p. 229, ma traduction ; voir aussi G. WILLIAMS 1988)[228]

Il faut mentionner que certains détracteurs de l'éthique évolutionniste, en visant à réduire au maximum l'impact de l'évolution sur notre activité morale, défendent souvent une position similaire (par exemple NAGEL 1983/1978) ; mais au contraire des éthiciens évolutionnistes, ils s'efforcent de réduire au maximum le nombre de ces capacités en utilisant des catégories très vagues comme l'intelligence, la raison ou le

[228] « Ethical systems (…) must have been produced indirectly by some sort of accident, the sort of thing that happens routinely in evolution. (…) These motivations must arise from biologically normal attitudes favored by kin selection and reciprocity, but have biologically abnormal manifestations in our abnormal modern environment. It is on this abnormality that human ethics depends. » (G. WILLIAMS 1993, p. 229).

libre arbitre et ne cherchent pas à produire une explication de l'évolution et de l'adaptation de ces capacités. A l'opposé, on trouve des défenseurs de l'éthique évolutionniste qui fragmentent et détaillent l'ensemble des capacités et tendances psychologiques sur lesquelles repose la moralité. Chandra SRIPADA et Stephen STICH (2006) par exemple défendent une théorie de la modularité massive (voir p. 42) combinée à l'idée que notre activité morale « surfe » sur un grand nombre de modules, c'est-à-dire de mécanismes qui ont évolué de manière propre. Parmi ceux-ci, il y a le mécanisme d'acquisition des normes, ou la tendance à adopter les normes et comportements de personnes prestigieuses. Ces auteurs pensent même que certaines tendances innées influencent le choix du contenu de nos normes ; c'est le cas par exemple du dégoût face à l'inceste. Entre ces deux extrêmes, on trouve toute une gamme d'auteurs (dont G. WILLIAMS, ROTTSCHAEFER ou PRINZ) qui défendent des positions intermédiaires.

Quoique Robert TRIVERS ne se prononce pas explicitement sur la question de savoir si la morale est un produit dérivé de l'évolution d'autres capacités, je pense qu'il pourrait également être classé dans cette position. Dans son fameux article de 1971, il produit une explication des mécanismes à l'origine de notre intelligence et de sentiments propres à la morale. Ils auraient évolué pour répondre aux problèmes adaptifs de nos ancêtres. TRIVERS part de la constatation que les êtres humains vivant en groupe d'individus qui interagissent régulièrement ont tout avantage à produire des chaînes d'interactions sur le modèle de l'altruisme réciproque. Mais du point de vue individuel, ils ont également avantage à profiter des bienfaits d'autrui sans contribuer en retour. Nous aurions donc en nous à la fois des tendances à l'altruisme et à l'opportunisme. La sélection naturelle se serait alors chargée de développer tout un système psychologique de plus en plus raffiné pour assurer le bon fonctionnement de la coopération en dépit des tricheries occasionnelles déclenchées par l'opportunisme. Elle nous aurait dotés de la capacité de former des amitiés, de ressentir de la gratitude ; elle aurait forgé des systèmes de détection des tricheurs, de désir de punir ces tricheurs, et inversement des formes très subtiles de malhonnêteté, d'hypocrisie ou de mensonge. Cette course à la tricherie et à la détection de la tricherie serait peut-être à l'origine de notre intelligence (ou en tout cas de la plus grande subtilité de nos capacités cognitives) ainsi que d'émotions typiques de la morale comme la culpabilité ou l'indignation.

La plupart des tenants de l'éthique évolutionniste optent cependant pour l'idée que la moralité elle-même serait une adaptation évolutionnaire ; elle ne serait pas simplement un produit dérivé. La moralité serait alors un ensemble assez compact et bien défini de dispositifs qui ont évolué précisément parce qu'ils permettaient de subvenir à un ensemble de besoins liés à la vie en communauté. Les divergences interviennent dans la conception précise du phénomène de la moralité ainsi que dans la détermination des processus qui ont mené à sa naissance et à sa stabilisation au fil de l'évolution. Ainsi la tâche que se donnent ces auteurs est de déterminer la fonction évolutionnaire (au sens étiologique du terme) propre à la moralité ; il s'agit de la définir en termes d'avantages sélectifs. La réponse à cette question dépend largement de notre connaissance des conditions de survie de nos ancêtres. Mais malheureusement, comme nos connaissances sur cette question sont assez maigres, les conjectures ont la part belle.

Se basant sur des considérations biologiques, anthropologiques et surtout sur les résultats de la théorie des jeux, certains auteurs pensent que la moralité est adaptative au niveau individuel et répond à un besoin, apparu au cours de l'évolution humaine, de régler les interactions entre les hommes.[229] Ce genre d'explications se décline en différentes variantes plus ou moins compatibles entre elles.

Une idée est que la moralité favorise la coopération qui, lorsqu'elle fonctionne bien, est à l'avantage de tout le monde. Selon Robert RICHARDS (1986, p. 289) par exemple, les sociétés humaines ancestrales étaient composées de petits groupes d'individus apparentés qui entraient régulièrement en compétition. Ce type d'environnement aurait été favorable à l'évolution de pulsions altruistes, lesquelles avaient pour fonction de servir le bien de la communauté. Ainsi un sens moral aurait évolué chez les êtres humains : un ensemble d'inclinations ou de dispositions naturelles qui engagent les individus à agir pour le bien de la communauté dont ils font partie. Plus précisément, cette attitude innée qu'est le sens moral est une attitude altruiste qui aurait évolué sous la pression de la sélection de parentèle et de la sélection de groupe dans le cadre d'un mode de vie en petites communautés. C'est donc grâce aux forces de la sélection de parentèle et de la sélection de groupe que les gens sont enclins à agir pour le bien de la communauté, c'est-à-dire de manière altruiste et donc morale.

[229] Michael RUSE par exemple écrit ceci : « Our moral sense, our altruistic nature, is an adaptation – a feature helping us in the struggle for existence and reproduction – no less than hands and eyes, teeth and feet. It is a cost-effective way of getting us to cooperate, which avoids both the pitfalls of blind action and the expense of a superbrain of pure rationality. » (1986, p. 99)

Michael RUSE (1984) propose une explication similaire mais au lieu de la sélection de parentèle ou de la sélection de groupe, il préfère donner de l'importance à l'idée de réciprocité élargie. Selon lui, le principe de la réciprocité serait ancré dans notre espèce et se manifesterait à notre conscience sous forme de sentiments moraux.[230] De manière générale, Ruse croit en l'existence d'un sens moral, un sens du bien, du mal et de l'obligation qui serait inscrit dans notre matériel génétique et se développerait au cours de notre ontogenèse. Plus précisément, il s'agirait d'un ensemble de dispositions[231] qui se manifestent sous forme d'émotions[232] lesquelles nous incitent à agir de manière altruiste[233] et nous insufflent la croyance en l'objectivité de nos convictions altruistes. Ainsi il écrit :

> « Je suggère que nous autres êtres humains, possédons une capacité innée (ou si l'on veut instinctive) de travailler ensemble de manière sociale. Et je suggère que cette capacité se présente au niveau physique sous forme d'un sens moral – comme un altruisme authentique de type Mère Teresa. Je défends donc – sur la base de considérations purement naturalistes, darwiniennes – que la moralité, ou plutôt le sens moral – une sensibilité à l'appel de l'altruisme et une propension à obéir – est câblé chez les êtres humains. Il est l'œuvre de la sélection naturelle pour nous pousser à travailler ensemble ou à coopérer. » (RUSE 2002, pp. 151-167, ma traduction)[234]

[230] « We feel that we ought to help others and to co-operate with them, because of the way that we are. That is the complete answer to the origins and status of morality. » (RUSE 1998, p. 252)

[231] « Mon idée est que nous avons des dispositions innées non pas simplement à être sociaux, mais aussi à être authentiquement moraux. » (RUSE 1993/1991, p. 52)

[232] RUSE distingue les sentiments moraux des sentiments traditionnels en affirmant que les premiers sont liés aux exigences de prescriptivité et d'universalité. « Here we start to move towards genuine morality and its evolution – from 'altruism' (in the biological sense of working harmoniously together, thus promoting reproductive ends), to altruism (in the literal sense, demanding genuine sentiments about right and wrong). » (RUSE 1998, p. 221)

[233] Plus précisément, RUSE pense que les émotions morales nous fournissent des motifs altruistes psychologiques qui nous pousseront à agir de manière altruiste (au sens évolutionnaire du terme). Ainsi, il ramène la moralité à l'altruisme.

[234] « I suggest that we humans have built in innately, or instinctively if you like, a capacity for working together socially. And I suggest that this capacity manifests itself at the physical level as a moral sense – as genuine, Mother Teresa-type altruism! Hence I argue – on purely naturalistic, Darwinian grounds – that morality, or rather a moral sense – a recognition of the call of altruism and a propensity to obey – is something which is hard wired into humans. It has been put there by natural selection in order to get us to work together socially or to cooperate. » (RUSE 2002, pp. 151-167)

De même que RICHARDS et RUSE, Larry ARNHART pense que les êtres humains possèdent un sens moral naturel (1998 ; 2000). Mais au contraire des précédents, ARNHART ne réduit par la moralité à l'altruisme. Pour lui, le sens moral consiste en un composé d'émotions morales (sympathie, culpabilité, indignation) et de principes moraux (sensibilité aux besoins d'autrui, tolérance, réciprocité) typiques des êtres humains. Sur l'échelle de l'évolution, le sens moral est le prolongement naturel du comportement pro-social. Il repose sur une propension à dispenser des bienfaits au-delà de la parenté directe et du conjoint ainsi que sur un ensemble de désirs partagés par tous les êtres humains (par exemple le désir réciproque des parents et de leurs enfants de rester ensemble ; 1998, p. 89). En bref, la fonction de la moralité consisterait dans le fait de faciliter les interactions positives entre individus.

Pour Christopher BOEHM (1997 ; 2002/2000), la moralité aurait évolué parce qu'elle propose une réponse adéquate à des situations de conflit à l'intérieur des groupes. Ainsi, BOEHM associe plus ou moins la moralité au contrôle social et à la régulation de conflits sociaux. Sa fonction serait d'éradiquer les comportements susceptibles de créer des conflits sociaux comme l'opportunisme ou l'influence démesurée des dominants, laquelle menace d'imposer un modèle social de dominance hiérarchique au détriment d'une certaine égalité entre membres de la société. La moralité serait née à partir du moment où, le langage aidant, les efforts individuels pour réduire les conflits à l'intérieur du groupe ont été développés à l'échelle de la collectivité (comportements de consolation, de réconciliation, de pacification active).

Richard ALEXANDER (1979, 1987, 1993) soutient que les systèmes moraux sont nés de la contradiction entre les intérêts individuels et les intérêts collectifs dans le cadre de conflits internes *et* entre groupes (à ce propos, voir aussi DE WAAL, 1997/1996, pp. 42-46). Deux niveaux sont ici à considérer. La sélection de groupe aiderait directement les individus qui font partie de groupes constitués d'individus eux-mêmes moraux ; en cas de conflit, un groupe composé d'individus soudés qui s'entraident augmente ses chances de victoire. A l'intérieur du groupe, il vaut la peine d'être un individu moral si la moralité est le prix à payer pour être admis dans le groupe (les êtres non moraux ont intérêt à agir de manière conforme à la morale s'ils ne veulent pas se faire rejeter par leurs voisins) et si elle assure un bon statut social (ce qui produit un effet personnel

bénéfique sur le long terme)[235]. Ce dernier point relatif à la réciprocité indirecte trouve écho chez Robert FRANK (1988) pour qui nos sentiments moraux ont évolué parce qu'ils sont indirectement avantageux pour les individus qui les produisent. Par exemple, les individus prêts à se sacrifier pour le bien-être d'autrui et disposés à payer de leur personne pour punir les opportunistes seront reconnus comme d'excellents partenaires de coopération. Dans le cadre de sociétés à fort contrôle social et faible mobilité, cette bonne réputation leur rendra service sur le long terme.

Au contraire de tous les auteurs (à l'exception partielle d'ALEXANDER) qui expliquent la moralité en termes d'avantages individuels, David Sloan WILSON (2002) propose une explication en termes d'adaptation au niveau du groupe.[236] Selon lui, appliquer des normes morales comme la règle d'or n'est pas avantageux du point de vue individuel mais au niveau du groupe. En cas de conflits entres groupes, un groupe composé d'individus immoraux perdrait contre un groupe composé d'individus moraux ; le second disposerait d'une meilleure cohésion sociale. Cette explication implique que la moralité ne peut être sélectionnée qu'au prix de guerres perpétuelles entre groupes (à ce propos, voir section 2.3.4). D. WILSON semble admettre ce point.

La liste des différentes explications de la genèse de la moralité n'est de loin pas close. D'autres auteurs combinent différents modèles présentés ci-dessus. Dennis KREBS (2002/2000) par exemple pense que la moralité a été sélectionnée à la fois parce qu'elle favorise la coordination et la coopération entre les individus d'une société et parce qu'elle permet d'établir une certaine égalité entre les membres d'une société (dans la même ligne, voir aussi DEHNER 1998). D'autres complexifient le tableau en faisant intervenir le niveau de la sélection culturelle (GINTIS 2003 ; LAHTI 2003).

Dans l'ensemble, toutes ces explications invoquent des mécanismes typiquement utilisés dans les théories évolutionnistes : sélection de parentèle élargie, réciprocité directe ou indirecte, sélection de groupe. Toutefois, étant donné le peu d'indices empiriques et historiques dont nous disposons, il n'est pas évident de choisir parmi ces

[235] ALEXANDER est un grand défenseur de la théorie de la réciprocité indirecte (section 2.3.3) ; par moments, il interprète même les systèmes moraux comme des systèmes de réciprocité indirecte.

[236] Pour ne pas trop caricaturer la position de David WILSON, précisons que lorsqu'il traite de la sélection de groupe, il a en tête les phénomènes et de sélection culturelle de groupe et de coévolution gène-culture.

options ; sans doute recèlent-elles toutes une part de vérité puisqu'elles réfèrent à différents aspects de la dynamique sociale dont la moralité fait indéniablement partie.

Le fait de concevoir la moralité en général comme une adaptation me paraît cependant hautement douteux. Il est possible de proposer des explications assez convaincantes de l'évolution d'un organe, d'une capacité ou d'une tendance comportementale. Mais peut-on vraiment parler de moralité en tant qu'objet de sélection ? Et si oui, de quel objet exactement s'agit-il ? Les auteurs ne sont pas très clairs sur ce point. Ce n'est certainement pas un organe, ni une simple tendance comportementale. Au mieux il pourrait s'agir d'une capacité ; par exemple la capacité d'établir une distinction entre ce qui est bien et ce qui est mal et à être motivé à agir en conséquence. RUSE par exemple dirait que nous avons la capacité de discerner les actions altruistes des actions non altruistes et cette compréhension cognitive déclenche en nous des émotions qui vont nous pousser à agir en sorte que le scénario altruiste se réalise. ARNHART dirait quant à lui que nous possédons des principes moraux et désirs innés qui orientent nos choix moraux. Mon doute face à ce genre de théories provient du fait que la moralité se voit réduite à des réactions intuitives sur lesquelles nous n'avons pas prise. Or il paraît évident que les codes moraux que nous formons en nos esprits sont le fruit de l'apprentissage et de nos réflexions sur des situations sociales conflictuelles. Cette dimension « particulariste » de la moralité (en particulier l'aspect réflexif) ne peut pas être expliquée par une approche évolutionnaire. Au fond, lorsqu'il est conçu en termes d'objet de sélection, le phénomène de la moralité est réduit à une réalité bien trop minimale (un simple ensemble de tendances comportementales et leurs mécanismes psychologiques proximaux) qui ne représente qu'une pâle image de ce que l'on comprend normalement par ce phénomène.[237] Je pense qu'il s'agit là d'une faiblesse significative de l'approche de la moralité comme adaptation. En bref, je suggère qu'il s'agit d'un phénomène bien trop complexe pour en faire un simple objet de sélection. Luc FAUCHER rejoint cet avis : « Mon point de vue est qu'il faut probablement éviter de considérer le domaine de la moralité comme s'il regroupait des phénomènes dont la structure commune profonde est identique, c'est-à-dire comme l'équivalent d'une espèce naturelle » (2007, p. 116).

[237] A ce propos, voir également la critique adressée contre RUSE par William ROTTSCHAEFER et David MARTINSEN 1990, p. 153-157.

Une autre objection que l'on pourrait faire à ce type d'explications[238] de l'évolution de la moralité est qu'elles commettent une erreur de méthodologie (à ce propos, voir COLLIER & STINGL 1993, p. 52 ; KITCHER 1987/1985, pp. 126-127): il s'agit de partir du principe que ce que l'on peut observer de nos jours est optimal avant de chercher des explications en termes d'adaptations plus anciennes.[239] Or il est évidemment faux de penser que tous les traits que l'on peut observer de nos jours sont optimaux ou même l'ont été ; un trait peut très bien être un produit dérivé qui n'a pas été sélectionné *pour* les effets qu'il produit (à ce propos, voir p. 165).

Face aux difficultés de la théorie de la moralité en tant qu'objet de sélection, les approches de la moralité en tant que produit dérivé semblent bien plus convaincantes. D'une part elles ne commettent pas l'erreur sélectionniste, d'autre part elles permettent une grande flexibilité explicative ; la moralité peut être conçue comme un phénomène complexe et chaque entité (capacité, tendance, émotion, etc.) sur laquelle on décide de la faire reposer est susceptible d'une explication évolutionnaire qui lui est propre.[240] C'est la voie que je me propose d'emprunter dans la suite de ce chapitre. Mais pour savoir sur quelles capacités et biais psychologiques repose la moralité, il faut en définir précisément les contours. Je procéderai en trois temps. Il s'agira d'abord de présenter un tableau, largement inspiré de données empiriques, de la manière dont je conçois l'activité évaluative et normative (section 5.2). Ce tableau ne s'avèrera cependant pas suffisant pour individuer la moralité. Une tentative sera alors faite en direction des émotions morales (section 5.3) ; elle se soldera cependant par un échec. En fin de compte (section 5.4) je proposerai deux critères suffisants pour délimiter le champ de l'activité morale. Cela me permettra de définir précisément les éléments sur lesquels elle repose.

[238] Cette objection est également couramment dirigée contre la psychologie évolutionniste.

[239] « Their arguments commonly start with observed behaviour, give a story about how such behaviour might enhance gene survival, and conclude that the behaviour is the consequence of a genetic adaptation expressed through epigenetic rules. » (COLLIER & STINGL 1993, p. 52)

[240] Précisons que cette conception de la moralité comme produit dérivé implique que la moralité ne possède aucune fonction (au sens étiologique du terme) qui lui est propre.

5.2. Un tableau affectif de l'activité évaluative et normative

Allan GIBBARD (2002/1990), un des rares philosophes de la morale qui prennent les données empiriques réellement au sérieux, propose de comprendre le phénomène de la moralité du point de vue psychologique. A cet effet, il a développé une explication détaillée de ce qui se passe chez les gens lorsqu'ils émettent des jugements moraux et forgent leurs convictions morales. Il a également largement contribué à éclairer les dynamiques des facteurs internes (tendances psychologiques) et externes (influence de l'entourage) dans la formation de nos évaluations et décisions morales. C'est ce type d'approche que je me propose d'adopter dans cette section. Même si je suis en désaccord sur bien des points avec l'analyse de GIBBARD, les connaisseurs de cet auteur remarqueront qu'il est le précurseur de beaucoup d'idées développées dans ce chapitre.

Grâce aux recherches menées depuis une vingtaine d'années à la fois en philosophie des émotions et dans les différentes branches des sciences cognitives (voir DE SOUSA 2003 ; première partie de ROBINSON 2005), il n'est plus gère possible de considérer l'activité évaluative sans prendre en compte les émotions. Il serait donc intéressant de disposer d'une explication de notre pensée et activité évaluatives qui précise non seulement le rôle et le fonctionnement des processus réflexifs mais également des émotions. Plusieurs explications de ce genre ont déjà été proposées. Pour ce qui est des jugements moraux, par exemple, certains ont avancé qu'ils consistent en des formes d'émotion (AYER 1946/1936), ou qu'ils portent sur des émotions (GIBBARD 2002/1990), ou qu'ils découlent d'expériences émotionnelles (TAPPOLET 2000, GOLDIE 2000). Le « tableau affectif »[241] présenté dans cette section propose une vue originale des rôles respectifs des processus cognitifs et affectifs dans l'activité évaluative en général (la question de l'activité morale en particulier sera réservée pour des sections ultérieures).

La présentation de ce tableau se fera en deux parties. Il s'agira d'abord de dégager les grandes lignes de l'activité évaluative et normative dans ce que j'appellerai le « tableau affectif de l'évaluation ». Ensuite sera traitée la question de la motivation à

[241] La notion de « tableau affectif » réfère au fait que l'analyse se situe au niveau purement descriptif (il s'agit de dépeindre de manière systématique la pensée et l'activité morales) et s'inscrit dans un courant de pensée qui accorde un rôle important aux émotions dans l'activité morale, au sens où elles sont étroitement liées aux jugements moraux et guident nos actions (GIBBARD 2002/1990, NICHOLS 2004).

l'action dans le « tableau affectif de la motivation ». Chacune de ces parties est largement inspirée de données empiriques auxquelles seront consacrées deux sections.

5.2.1. Le tableau affectif de l'évaluation

Avant d'exposer les détails du tableau affectif, je propose de passer en revue quelques données empiriques dont il s'inspire.

i. Quelques données empiriques relatives aux jugements moraux

De manière générale, ce que nous montre un nombre grandissant d'études menées en psychologie est que les gens font souvent des jugements largement automatiques et non réflexifs. Une manière éclairante d'interpréter ces résultats est de comprendre nos jugements comme le résultat de simples réactions émotionnelles. Prenons par exemple le cas de l'inceste. Dans une étude très connue menée par Jonathan HAIDT (2001) montre que la plupart des gens condamnent de manière impulsive les pratiques incestueuses et soutiennent ce verdict même au terme d'une discussion où on les force à admettre qu'ils ne disposent d'aucune bonne raison pour fonder leur jugement.

Plus récemment, Thalia WHEATLEY et Jonathan HAIDT (2005) ont mené une expérience sur des sujets hautement hypnotisables. Sous condition d'hypnose, ils ont expliqué aux sujets qu'ils éprouveraient du dégoût chaque fois qu'ils liraient un mot arbitraire (par exemple le mot « donc »). Une fois réveillés, ils ont demandé aux sujets de lire et juger moralement une série de petites histoires assez communes (dont certaines n'étaient même pas particulièrement pertinentes au plan moral) qui contenaient ou non le mot lié au dégoût hypnotique. Les résultats sont impressionnants : systématiquement, les sujets condamnent moralement les histoires contenant les mots qui causent en eux du dégoût. Cette étude montre que des émotions causées de manière artificielle sont également génératrices de réactions morales.

Dans une expérience congruente menée par GREENE et collègues (2001), les sujets doivent mener des expériences de pensée sur différentes variantes de situations problématiques comme celle du trolleybus où il s'agit de choisir de tuer une personne

afin d'éviter la mort de cinq autres personnes.[242] Au cours de l'expérience, on demande aux sujets quelle action, parmi deux options données, ils considèrent comme moralement adéquate. Dans le cadre de l'expérience, le cerveau des sujets est scanné à l'aide de la technique d'imagerie cérébrale. Les résultats semblent montrer que l'engagement émotionnel influence largement les jugements moraux : s'imaginer devoir pousser une personne sous un trolleybus roulant à pleine vitesse afin de le stopper, et par là, sauver les cinq passagers, est émotionnellement plus saillant que s'imaginer pousser une manette qui va orienter la trajectoire du trolleybus sur une voie où se trouve un homme – dans cette deuxième option, la vie de l'homme sera également sacrifiée pour stopper le trolleybus. Cette différence d'engagement émotionnel induit les sujets à condamner la première action et à juger la seconde comme moralement permissible, alors même que dans les deux cas, la vie d'une personne est sacrifiée pour sauver celle de cinq autres.

Cette interprétation se trouve confirmée par une expérience similaire menée avec des sujets normaux et des patients souffrant de lésions du cortex frontal et ventromedial, des parties du cerveau dont on sait qu'elles sont liées aux expériences affectives. Malgré leurs déficiences au niveau affectif, les capacités intellectuelles des patients étaient préservées et ils se montraient capables de distinguer entre les situations moralement et non-moralement pertinentes. Toutefois, les tests ont montré que leurs jugements au sujet des situations morales étaient biaisés par rapports aux sujets normaux : au contraire de ces derniers, les patients considéraient les deux scénarios du trolley mentionnés plus haut comme également acceptables. Tout porte à croire que la cause de cette absence de discernement est due à la déficience de leur système affectif (CIARAMELLI *et al.* 2007).

Enfin, dans différentes études, des psychologues ont testé la manière dont les gens justifient leurs jugements moraux (MIKAIL 2002 ; HAUSER 2006 ; CUSHMAN *et al.* 2006). On présente aux sujets des problèmes moraux comme celui du trolleybus. Puis, on leur demande dans un premier temps quelles actions ils considèrent comme moralement acceptables ou moralement condamnables et dans un deuxième temps comment ils justifient leur jugement. L'analyse des réponses données par les sujets suggère l'existence de certaines règles intuitives profondément ancrées dans leur esprit, qui guident leurs jugements même si elles ne font pas surface dans leurs raisonnements conscients ; lorsqu'il s'agit de justifier leurs évaluations spontanées, les sujets

[242] Le premier modèle de cette expérience de pensée est dû à la philosophe Philippa FOOT (1967).

n'invoquent pas toujours les règles qui les ont réellement poussés à formuler ces jugements.[243] Parmi ces règles intuitives, il y aurait celle selon laquelle un tort (*harm*) causé de manière intentionnelle en tant que moyen pour atteindre une fin est moralement plus condamnable que le même tort s'il est conçu comme l'effet secondaire d'un but que l'on cherche à atteindre.[244] Comme autre exemple, les gens conçoivent un tort causé directement avec contact physique comme moralement plus condamnable que le même tort causé sans contact physique.

Notons que ce dernier groupe de données ne nous dit rien sur le rôle des émotions dans les jugements moraux. Ces auteurs s'expriment même de manière plus ou moins explicite contre l'idée d'un rôle prépondérant assumé par les émotions. Dans leur article, MIKAIL et collègues par exemple, critiquent l'idée de GREENE et collègues selon laquelle l'engagement émotionnel influence le jugement moral (cf. expérience présentée plus haut). Je pense qu'il faut sérieusement remettre en doute la pertinence de cette critique car en observant de plus près les données obtenues par les deux groupes de recherche, on constate qu'elles sont parfaitement compatibles avec une interprétation qui fait dépendre nos jugements moraux conscients de la puissance des réactions émotionnelles. Il faut comprendre qu'une réaction émotionnelle est déjà une forme d'évaluation (ce point deviendra plus clair à la section iii). Ensuite, de manière générale, il me semble que tous les résultats obtenus dans le cadre des expériences sur la réaction des sujets face à des problèmes moraux tels que celui du trolley peuvent être expliqués au moyen de la notion de « réactions émotionnelles différenciées » sans faire appel à l'idée de « règles intuitives » : une classe particulière d'états mentaux (par exemple la prise de conscience qu'une souffrance est causée de manière intentionnelle) déclenche une réaction émotionnelle typique plus forte qu'une autre classe d'états mentaux (par exemple la prise de conscience qu'une souffrance est causée de manière non intentionnelle). La puissance de nos réactions émotionnelles est ensuite causalement responsable de nos évaluations morales conscientes (pour une analyse similaire, voir HAIDT & JOSEPH 2004).

[243] Pour des résultats similaires, voir aussi les travaux de BARON (1998) et de NISBETT & T. WILSON (1977).

[244] Il s'agit en fait de la doctrine du double effet qui avait déjà été introduite par Thomas D'AQUIN 1985/1265-1273, Partie II/II Qu. 64, Art.7).

ii. Les grandes lignes du tableau affectif

Les données empiriques présentées dans la section précédente soutiennent l'idée que les jugements moraux ne résultent généralement pas de processus d'inférence au cours desquels nous appliquons consciemment une norme ou une valeur[245] à une situation; il semblerait plutôt que les évaluations soient intuitives, émotionnelles, largement automatisées (alors même que certaines situations qui les déclenchent sont hautement subtiles).

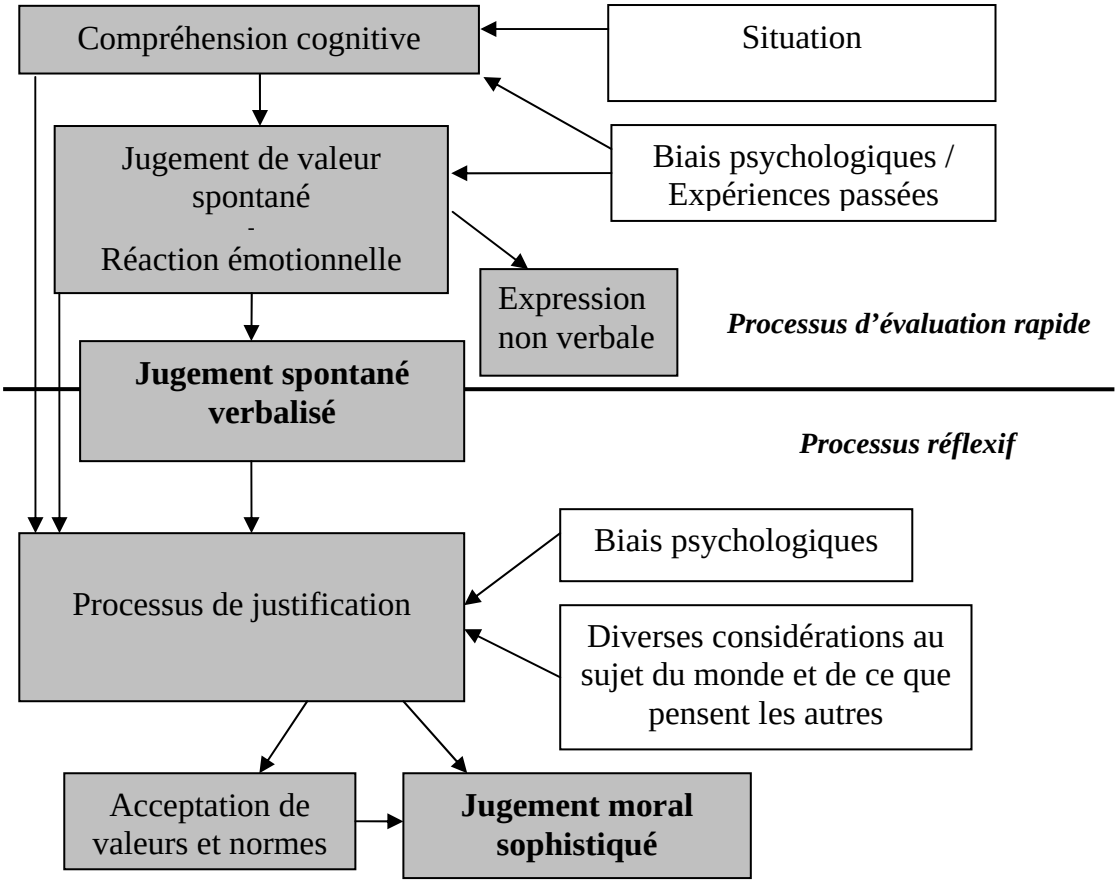
A première vue, on pourrait donc penser que les jugements moraux se réduisent à des expressions de réactions émotionnelles. Dans ce cas, on pourrait par exemple défendre une position comme l'émotivisme selon laquelle les jugements moraux sont de simples expressions d'approbation ou de désapprobation (AYER 1946/1936).[246] Mais je pense que l'image qui se dégage de ces études et de notre expérience courante de l'activité morale se doit d'être plus raffinée. Il faut rendre justice à notre capacité de mener des raisonnements complexes et, par le moyen de processus d'inférence et sans l'aide des émotions, de tirer des conclusions au sujet de ce qu'il faut faire. Même les psychologues comme Jonathan HAIDT (2001) ou Joshua GREENE (GREENE & HAIDT 2002) admettent cette réalité. Il est donc erroné de postuler que les jugements moraux ne sont rien de plus que des émotions ou des expressions d'émotions. Dans ce qui suit, je vais tenter de développer un tableau, largement inspiré des données empiriques présentées ci-dessus, qui précise la manière dont les gens évaluent les situations. Ce tableau portera cependant sur l'activité évaluative en général. La question extrêmement complexe et délicate de la délimitation du champ de la moralité fera l'objet des sections 5.3 et 5.4.

Pour commencer, voici un schéma grossier du tableau affectif de l'évaluation qui sera explicité dans les sections suivantes.

---

[245] Une définition plus précise de ces deux notions sera proposée à la section 5.4.2.

[246] A première vue, on pourrait également penser aux positions expressivistes à la Allan GIBBARD (2002/1990). Mais en fait, la théorie de GIBBARD est plus complexe. Elle n'affirme pas que les jugements moraux sont de simples expressions de réactions émotionnelles ; elle propose de les comprendre comme des expressions d'états d'esprit complexes qui consistent dans le fait de penser qu'il est justifié de ressentir une certaine émotion face à un certain type de situation.



Les flèches représentent des relations causales. Les boîtes grises représentent les différentes activités des agents et les boîtes blanches mentionnent ce qui influence l'activité des agents.

En deux mots, les grandes lignes de mon tableau affectif sont les suivantes : il y a deux processus dans l'activité évaluative. Le premier, que j'appellerai « processus d'évaluation rapide », est émotionnel et automatique alors que le second, que j'appellerai « processus réflexif », plus raffiné, inclut les réflexions conscientes et raisonnements d'inférence. En présentant ce double modèle, je montrerai qu'il existe différents types de jugements, certains plus spontanés de d'autres.

iii. Le processus d'évaluation rapide

Commençons par le premier processus. Lorsque nous sommes témoin de certaines situations nous formons des « jugements de valeur spontanés ». Ces jugements sont des sortes de réactions émotionnelles composées de deux parties intrinsèquement liées l'une

à l'autre : une évaluation rapide non inférentielle[247] d'un état de fait, et un état affectif particulier lié à une sensation particulière. En utilisant le terme de « sensation », je voudrais exprimer l'idée que les réactions émotionnelles possèdent une certaine phénoménologie dans notre expérience personnelle qui est intimement liée à l'évaluation.[248] Par son aspect évaluatif, la réaction émotionnelle possède un contenu intentionel (au sens où elle est dirigée vers un objet).[249] Elle est une manière de prêter une attention sélective à certaines caractéristiques d'une situation et de les percevoir d'une certaine manière (voir DE SOUSA 1987)[250] ; pour ce qui est des jugements de valeur spontanés, c'est une manière d'interpréter une situation comme « devant être réalisée » ou « devant être évitée » et cette évaluation s'exprime par un état affectif qui implique des sensations négatives ou positives (comme celles liées au fait de se sentir dégoûté ou de se sentir élevé)[251].

Cette manière d'interpréter le monde dépend de différents facteurs : certaines tendances génétiquement innées, nos expériences passées et l'environnement social dans lequel nous nous trouvons. Voyons dans le détail comment cela fonctionne.

---

[247] Pour une analyse du caractère non inférentiel de l'évaluation émotionnelle, voir Sabine DÖRING (2007).

[248] En philosophie des émotions, il y actuellement un grand débat autour de la question de savoir si l'aspect affectif de la réaction émotionnelle est causé par l'aspect évaluatif (ROBINSON 2005) ou si les deux aspects se confondent en un tout indissociable (GOLDIE 2000 ; DÖRING 2007). La première interprétation comprend l'aspect affectif en termes de « sensations physiques » (*bodily feelings*) qui réfèrent à l'idée jamesienne de prise de conscience d'un état corporel interne tel qu'une réaction musculaire ou un changement hormonal (JAMES 1884) ; les défenseurs de la seconde interprétation préfèrent comprendre l'état affectif comme une « sensation de l'émotion » (*emotion's feelings*) qui réfère à la notion de « sensation envers » (*feeling towards*) développée par Peter GOLDIE (2000). Cette dernière est intentionelle (au sens où elle porte sur un objet du monde extérieur) et indissolublement liée à l'évaluation contenue dans l'émotion. A la section 3.3.3, j'ai adopté une approche à la GOLDIE et DÖRING mais dans le cadre qui nous concerne ici, il n'est pas nécessaire de prendre position dans ce débat. Il est suffisant de dire qu'une réaction émotionnelle est à la fois affective et évaluative.

[249] Dans la mesure où les réactions émotionnelles ont un contenu intentionel, on peut dire qu'elles sont cognitives. Toutefois, cela n'implique pas que les réactions émotionnelles se réduisent à des jugements propositionnels, c'est-à-dire à des sortes de croyances (à ce propos, voir DÖRING, 2007).

[250] C'est une manière primitive et non réflexive de voir quelque chose comme susceptible de déclencher une émotion (colère, culpablitité, etc.) et d'être affecté en ce sens.

[251] L'élévation, un sentiment positif et chaleureux ressenti lorsque nous sommes témoin d'actes que nous considérons comme moralement bons, nous incite à aider autrui (HAIDT 2000).

Dans une certaine mesure, nos réactions émotionnelles sont régies par des valeurs simples et intuitives dont les résultats, après réflexion, nous satisfont généralement assez bien (quoique ce ne soit pas toujours le cas). Grâce à ces raccourcis mentaux (certains ne sont autres que les émotions altruistes dont il était question au chapitre 3), nous avons plus ou moins les mêmes réactions émotionnelles face à des situations similaires. Par exemple, nous sommes généralement plus choqués par les actions qui causent un tort à autrui que par les omissions qui causent un tort à autrui ;[252] de manière générale, nous supportons mal les actions qui causent de la souffrance chez autrui (NICHOLS 2004) de même que les comportements non loyaux ou opportunistes. Certains auteurs parlent de grammaire évaluative universelle,[253] d'un ensemble de mécanismes et de capacités dont la fonction est de générer des jugements rapides au sujet de ce qui est acceptable ou non (HAUSER 2006, MIKAIL 2002 ; HARMAN 1999). Au contraire de ces auteurs, je n'irai pas jusqu'à défendre cette idée de *grammaire morale universelle*. Premièrement, plutôt que des *règles* d'évaluation (qui seraient typiques d'une *grammaire*), je pense que ce qui est ancré dans nos esprits, ce sont des modèles de réactions émotionnelles face à certains types de situations : je les appellerai « valeurs intuitives ». Il se peut que ces modèles donnent l'impression de former des règles ; c'est le cas lorsque des réactions émotionnelles différenciées sont systématiquement causées par des situations similaires mais qui se distinguent par un critère crucial (par exemple, lorsque nous sommes plus fortement choqués par les *actions* qui causent un tort à autrui que par les *omissions* qui causent le même tort à autrui). Deuxièmement, je pense qu'une réaction émotionnelle n'est pas suffisante pour former un jugement *moral.* La raison en deviendra claire aux sections 5.2.1.v et 5.4. Troisièmement, parler d'*universalité* dans ce contexte me paraît exagéré (à ce propos, voir p. 175). Nos facultés émotionnelles sont plastiques et ouvertes, au sens où elles se développent en fonction de nouvelles expériences. Cela signifie qu'elles sont influencées par nos expériences passées et par le contexte dans lequel nous nous trouvons. En d'autres termes, toute réaction émotionnelle est particulière, car imbriquée dans la vie d'un sujet

[252] Notons en passant que nous réagissons de cette manière même dans les cas où l'action causerait moins de dommages que l'omission (à ce propos, voir BARON 1998 ; SUNSTEIN 2005).

[253] Dans le domaine du langage, CHOMSKY (1964) a développé l'idée que chaque être humain possède une « grammaire universelle », c'est-à-dire un ensemble de mécanismes innés qui fournissent un cadre général sur la base duquel se forme tout langage. Il se pourrait qu'une « grammaire » similaire existe en matière d'évaluation.

avec toutes ses contingences culturelles et personnelles. Considérons un exemple. Martial est grondé plusieurs fois par ses parents pour avoir caché les jouets de sa petite sœur. Disons que ces reproches l'ont mis mal à l'aise. Suite à cela, Martial associe de manière inconsciente le fait de subir des reproches avec le fait de se sentir mal à l'aise. Et puisque cette sensation est déplaisante, son cerveau marquera, au sens de la théorie des marqueurs somatiques de DAMASIO (2001/1994), les reproches ainsi que les situations susceptibles de les déclencher comme étant « à éviter ». Dans ce contexte, on peut parler de mécanismes appris (par opposition aux mécanismes purement innés) qui déclenchent une sensation émotionnelle qui à son tour guide les futures évaluations, choix et comportements. Dans notre exemple, chaque fois que Martial éprouve le désir d'une action susceptible de déclencher le reproche de ses parents (par exemple, cacher les jouets de sa sœur), son cerveau reconstituera le marqueur somatique du malaise (même si c'est de manière moins vive que dans les circonstances réelles) et cela le motivera à se restreindre d'agir de la sorte.[254]

Nos jugements de valeur spontanés dépendent également des divers motifs que nous avons par ailleurs. Par exemple, des recherches en psychologie sociale ont montré que les gens sont profondément influencés par le désir de maintenir des relations sociales plaisantes. Ce motif guide leurs attitudes évaluatives et la manière dont ils traitent l'information (CHEN *et al.* 1996). De manière plus générale, beaucoup de recherches ont également été menées sur le phénomène de la contagion émotionnelle, la tendance à ressentir des émotions similaires à celles d'autrui (SIMNER 1971 ; HATFIELD *et al.* 1994). En résumé, notre environnement social dirige nos réactions émotionnelles dans le sens de la consistance avec celles de nos voisins.

Pour terminer, précisons que même si nous sommes passifs dans nos réactions émotionnelles (au sens où elles s'imposent à notre conscience de manière soudaine, incontrôlée et sans effort), cela n'empêche pas que ces réactions émotionnelles puissent découler de pensées et croyances complexes ; elles peuvent résulter de la prise de conscience de spécificités très fines d'une situation. De plus, leur grain particulier dépendra en bonne partie du contexte culturel dans lequel nous évoluons, car les émotions complexes sont des choses que nous apprenons et qui sont largement forgées par la culture.

[254] De plus, comme nous le verrons plus loin, au cours du processus réflexif, Martial attribuera probablement une valeur négative à ce type d'actions.

iv. L'expression d'un jugement de valeur spontané

Un jugement de valeur spontané est généralement exprimé ; il ne reste pas tacite. Il y a deux manières de le faire. La première est de l'exprimer sous forme non verbale. C'est le cas par exemple lorsque notre faciès se contracte et exprime la colère, ou lorsqu'on crie « Ah ! » (AYER 1946/1936). Mais cette manière d'exprimer un jugement de valeur spontané reste encore dans le domaine de l'incontrôlable ; l'expression n'est que l'extension automatique et directe de la réaction émotionnelle interne. Elle s'inscrit dans une polarité positive (approbation) ou négative (désapprobation).

La deuxième manière d'exprimer un jugement de valeur spontané comporte une expression verbale. C'est-à-dire que l'on peut conceptualiser les réactions émotionnelles en termes de « c'est horrible ! » ou « c'est bien ! ». Il ne s'agit pas uniquement d'une expression d'approbation ou de désapprobation (comme l'expression non verbale) mais d'un énoncé conscient et réflexif par le moyen duquel on attribue une valeur à quelque chose. Il semblerait donc qu'il s'agisse d'une sorte de reconstruction cognitive de la réaction émotionnelle. Cet apport cognitif, qui s'ajoute à la forme primitive du jugement de valeur, permet en fait la production d'un nouveau jugement légèrement plus complexe : ce que je vais appeler un « jugement spontané verbalisé ». L'idée est qu'en mettant des mots sur un jugement de valeur spontané, nous ne nous contentons pas de l'exprimer mais produisons en fait un nouveau jugement.[255] De cette manière, nous faisons un premier pas en direction d'un processus réflexif dans le cadre duquel il s'agira de conceptualiser nos réactions émotionnelles en termes de valeurs et de normes qui seront ensuite appliquées à la situation.

Toutefois, même s'ils sont légèrement plus raffinés (car nous ne nous contentons pas d'exprimer une attitude mais attribuons une valeur à une situation), ces jugements verbalisés demeurent très confus et s'effectuent également de manière largement automatique. En produisant un jugement verbalisé, nous affirmons sans plus de précision que quelque chose est bon ou mauvais parce que nous le ressentons de cette

[255] Il semblerait que ce soit ce genre de jugements que les psychologues comme HAIDT ou HAUSER testent dans leurs expériences.

manière.[256] Sur une échelle de complexité, les jugements spontanés verbalisés se situent donc entre deux extrêmes : la simple expression d'approbation ou de désapprobation et l'assertion d'un énoncé normatif à caractère objectif.

v. Processus réflexif et mécanismes d'influence mutuelle

Venons-en à la partie la plus rationnelle ou cognitive de l'activité évaluative ; c'est ici qu'entrent en jeu les « jugements sophistiqués ». Ils peuvent être *évaluatifs* s'ils découlent uniquement de valeurs[257] ou *normatifs* s'ils s'appuient sur des normes lesquelles contiennent des valeurs[258]. La différence entre les jugements de valeur et les jugements normatifs tient à ce que les premiers ne sont pas directement liés au prescriptif alors que les seconds le sont (puisqu'ils découlent de normes, lesquelles sont prescriptives).

Nous nous trouvons souvent confrontés à des désaccords entre nos jugements de valeur spontanés et ceux produits par autrui ou entre nos jugements et le comportement d'autrui. Parce que notre survie dépend largement de notre capacité de mener une vie cohérente et coordonnée à celle de nos voisins, nous éprouvons un double besoin : d'un côté, nous voulons nous prouver à nous-mêmes et à autrui la pertinence de nos réactions émotionnelles, de l'autre côté nous désirons qu'autrui partage nos réactions émotionnelles. Pour réaliser ce double objectif, nous nous engageons dans une activité complexe au cours de laquelle nous réfléchissons sur les raisons qui justifient nos réactions émotionnelles et jugements spontanés verbalisés (c'est ce que j'appellerai le « processus réflexif ») et mettons en pratique de manière plus ou moins consciente diverses méthodes pour influencer les réactions et convictions de nos voisins.

Dans le cadre du processus réflexif, nous cherchons à justifier nos jugements spontanés. La justification, un processus qui requiert l'activité de la raison, est utilisé à

[256] A ce niveau de mon explication, je ne voudrais pas défendre une position volontariste qui accorderait un rôle trop important à la volonté dans la production des jugements spontanés verbalisés.

[257] Voici deux exemples de jugements de valeur non prescriptifs : on peut juger que Jésus est un héros parce qu'il a donné sa vie pour sauver son peuple, sans pour autant prescrire d'agir de la sorte ; de même, on peut juger qu'il est bon d'être compatissant face au malheur d'autrui sans qu'aucune prescription particulière ne soit impliquée dans ce jugement.

[258] Pour des exemples détaillés de jugements normatifs, voir section 5.4.3.

la fois pour convaincre une personne rationnelle du bien-fondé de l'objet que nous cherchons à justifier et pour tenter de rallier cette personne à notre cause.[259] Nous pouvons le faire de manière minimale en nous demandant simplement s'ils sont appropriés à la situation ; il s'agit alors de savoir si tous les aspects pertinents ont été pris en compte, si notre engagement personnel a altéré notre jugement, etc. Nous pouvons également chercher à leur donner une justification forte au sens de *fondement* ; dans ce cas, nous faisons reposer nos jugements sur des normes et des valeurs. Dès lors, pour les besoins de notre tableau, il serait intéressant de disposer d'une explication de la manière dont nous définissons et choisissons les normes et valeurs auxquelles nous adhérons.

Mon idée est que nous choisissons nos normes et valeurs en grande partie en fonction des sensations qui sont causées en nous dans des situations concrètes. Si un état de choses cause en nous une forte sensation désagréable, dès que nous y réfléchissons, nous aurons tendance à attribuer une valeur négative à cet état de choses (et inversement pour les sensations positives). Ce lien peut s'établir de deux façons : la première met en jeu les sensations qui font partie d'une réaction émotionnelle alors que la seconde met en jeu des sensations simples.

Les réactions émotionnelles, c'est-à-dire les jugements de valeur spontanés, peuvent révéler ce que nous valorisons inconsciemment. En effet, il semble qu'une partie de ce que nous valorisons nous est rendu épistémologiquement accessible précisément par la médiation des réactions émotionnelles (voir GOLDIE 2000, pp. 48-49). Par exemple, si nous sommes horrifiés à la vue de notre voisin qui bat son enfant pour le plaisir (et par là produisons un jugement de valeur spontané), nous aurons tendance à attribuer une valeur négative à ce type de comportement et à établir une norme correspondante qui interdit ce genre d'action. Selon cette explication, notre esprit est déjà imprégné de manière inconsciente par un certain nombre de valeurs intuitives et les réactions émotionnelles servent à nous en faire prendre conscience.[260] Toutefois, ce n'est certainement pas la seule manière de définir nos valeurs conscientes. Si c'était le cas, nous naîtrions avec un dispositif de valeurs intuitives prédéterminé et notre seul horizon serait de les découvrir.

---

[259] Pour une explication plus détaillée de la notion de justification, voir p. 245.

[260] A ce propos, voir aussi ROZIN *et al.* 1999; selon ces auteurs, certaines émotions sont intrinsèquement liées à certains types de codes moraux que l'on rencontre dans toutes les sociétés humaines.

Le choix d'une valeur peut aussi résulter d'une simple sensation associée à une représentation conceptuelle de la cause de cette sensation. Souvenons-nous de Martial ; après s'être fait gronder par ses parents, il éprouve un certain malaise chaque fois qu'il conçoit l'idée de cacher les jouets de sa sœur. Cette sensation de malaise associée à un certain type de situation l'incitera à valoriser négativement la situation en question ; il pensera que c'est mal de cacher les affaires de sa sœur. En bref, nous avons tendance à attribuer une valeur négative à ce qui cause en nous des sensations désagréables (et inversement pour les valeurs positives).[261] De plus, les valeurs et normes que nous choisissons sous l'influence de simples sensations peuvent même supplanter des valeurs et normes induites par des réactions émotionnelles préalablement ancrées en nous. Prenons un exemple parlant. Prosper prône la fidélité à tel point qu'il est dégoûté à l'idée même d'un individu qui trompe son partenaire. Mais un jour il rencontre Célestine et craque… il trompe sa femme avec elle. Cette expérience s'avère si plaisante qu'il cesse d'éprouver de l'aversion envers les partenaires infidèles.

L'exemple de Prosper et Célestine illustre également un autre principe : la puissance des sensations joue un rôle dans la manière dont nous choisissions nos normes et valeurs ; plus une sensation est forte, plus nous avons tendance à attribuer une valeur à ce qui en est la cause.

Voilà pour une première explication de la manière la plus courante de choisir nos normes et valeurs. Il existe cependant d'autres manières de les définir. Nous pouvons par exemple nous lancer dans des raisonnements d'inférence plus complexes et, partant de normes et valeurs préalablement acceptées, en déduire de nouvelles. C'est à ce stade que la pensée rationnelle prend sa place dans l'activité évaluative. Toutefois, sans parler du fait que tout processus d'inférence part de prémisses données, il faudrait se garder d'y accorder trop d'importance. Comme le font bien remarquer Joshua GREENE et Jonathan HAIDT (2002), il est vrai que les gens s'engagent souvent dans des débats moraux réels ou fictionnels mais la plupart du temps, ces efforts sont dirigés vers le

[261] Sabine DÖRING (2007) défend un point de vue similaire mais basé sur une compréhension des émotions comme perceptions. Selon elle, « en faisant l'expérience d'une émotion, le monde apparaît au sujet comme s'il était tel que l'émotion le lui représente ». Elle ajoute que « les émotions jouent un rôle dans le raisonnement avant que le raisonnement ne joue son rôle dans la rationalisation d'une action » – cela dit, comme nous le verrons plus loin, je m'éloigne de la position de DÖRING lorsqu'elle prétend que les émotions permettent de percevoir des valeurs extérieures.

renforcement ou la transmission de jugements, valeurs ou normes sur lesquels les sujets ont déjà fixé leur choix par avance.

Dans les contextes sociaux, les gens tentent de s'influencer mutuellement et d'assurer un consensus avec leurs amis ou alliés. Il existe même des mécanismes psychologiques (auxquels est consacrée toute une littérature empirique) qui régissent et assurent l'efficacité de cette influence mutuelle. Selon les anthropologues évolutionnistes Joseph HENRICH, Robert BOYD et collègues, les êtres humains sont largement influencés dans leurs choix, jugements et pratiques par un certain nombre de biais psychologiques. Les biais sont des sortes de règles d'apprentissage social du type « Copie qui a le plus de succès ! », ou « Copie la majorité ! » Comme nous l'avons vu dans la section 1.2.3 (p. 43), un des biais les plus influents est celui du conformisme : les êtres humains ont tendance à reproduire les comportements les plus fréquents de la population dans laquelle ils évoluent (HENRICH & BOYD 1998). Un autre biais est celui du prestige : les êtres humains tendent à prendre pour modèle des individus qui paraissent avoir du succès ou qui semblent posséder des qualités ou des connaissances supérieures (HENRICH & GIL-WHITE 2001). Ces mécanismes psychologiques ont probablement évolué parce qu'ils permettent aux individus de bénéficier à peu de frais des avantages liés à l'adoption d'un comportement, d'une norme, d'une valeur ou d'une coutume. Chandra SRIPADA et Stephen STICH (2005, pp. 150-155 ; voir aussi FESSLER & NAVARRETE 2003) renforcent à l'aide d'exemples concrets cette hypothèse de l'existence de biais psychologiques.[262] Faisant référence à différentes études empiriques, ils montrent que les êtres humains sont souvent incapables de juger correctement les avantages et désavantages de variantes culturelles (par exemple dans le domaine des innovations ou des tabous liés à la nourriture) et dirigent leurs choix en suivant les biais du conformisme et du prestige.[263] De plus il a été montré dans différentes expériences psychologiques menées par Jody DAVIS et Caryl RUSBULT

[262] Notons qu'en 1990 déjà, Allan GIBBARD intégrait dans son explication de l'activité morale, des mécanismes psychologiques comme l'« influence normative » (*normative influence*), similaire au biais du prestige, ou comme l'exigence de cohérence (*demand for consistency*), similaire à l'émotion épistémique qui sera introduite plus loin dans cette section.

[263] Ils soulignent également le fait que, même s'ils sont souvent des révélateurs d'adaptations, ces biais peuvent mener à la propagation d'innovations et pratiques hautement maladaptives si elles sont prises isolément (c'est par exemple souvent le cas en matière de tabous alimentaires).

(2001) que nous sommes directement influencés dans nos jugements par ceux de nos amis, alliés ou proches parents. Cela confirme selon eux l'existence du phénomène d'« alignement d'attitude » (*attitude alignment*), une tendance, chez les partenaires d'interaction, à modifier leurs attitudes respectives de manière à ce qu'elles convergent (pour plus de données, voir HAIDT 2001).[264]

En bref, nous formons essentiellement des valeurs et normes congruentes avec nos sentiments et les choix de nos voisins. Cela n'empêche en rien cependant l'émergence de conflits à l'intérieur même de nos productions normatives (c'est-à-dire nos actions, les valeurs ou normes auxquelles nous souscrivons, les jugements spontanés ou sophistiqués que nous produisons) ou entre nos productions normatives et celles de nos voisins. C'est ici que l'activité rationnelle refait surface dans le tableau affectif : c'est elle qui attire notre attention sur ce genre d'incohérences. Cette prise de conscience déclenche en nous une émotion épistémique[265] que je nommerai « demande de cohérence », qui se caractérise par un sentiment d'inconfort. En effet, promouvoir à la fois p et –p, nous met dans une situation analogue à un dilemme pratique où il est impossible de réaliser tous les buts que nous nous sommes fixés. Ce sentiment nous inclinera à tenter de rétablir la cohérence (pour des études empiriques supportant cette idée, voir HAIDT 2001 ; MOSKOWITZ *et al.* 1999 ; THAGARD 1992). Nous ne pourrons pas nous empêcher de procéder à des raisonnements inférentiels, à ouvrir notre cœur à nos sentiments les plus profonds et à chercher l'avis de notre entourage jusqu'à ce que l'harmonie soit rétablie. Enfin, l'activité rationnelle signifie également la prise en considération des questions d'ordre pratique et l'orientation du choix de nos normes en fonction de leur réalisabilité et des autres buts que nous nous sommes fixés.

[264] Précisons également que les mécanismes psychologiques mentionnés sont étroitement liés aux émotions. Par exemple, le biais du prestige se manifeste par une certaine sensibilité émotionnelle face à des personnes que l'on considère comme prestigieuses ; il induit la tendance à copier les pratiques et adopter leurs croyances et valeurs.

[265] Certains préfèreront parler de mécanisme psychologique (GIBBARD 2002/1990).

vi. Objectivité réelle et objectivité psychologique

Etant donné que le tableau affectif fait grand cas de la manière dont les gens forgent leurs jugements évaluatifs au cours de leur ontogenèse et au fil de leurs expériences quotidiennes, il rend parfaitement compte de la diversité évaluative et normative que l'on peut constater à l'intérieur des cultures et surtout entre les cultures (OKLESHEN 1996 ; HENRICH *et al.* 2004 ; PRINZ 2007). Toutefois, il est bien connu qu'il y a aussi un bon nombre de similitudes ; la plupart des sociétés humaines condamnent le viol, le meurtre, l'inceste, le vol ou le mensonge et louent l'honnêteté, le partage, l'aide et la réciprocité (CASHDAN 1989 ; BOEHM 1999). Dans cette section, j'aimerais montrer que ce phénomène est également compatible avec le tableau affectif. Nous verrons que ce dernier fait ressortir deux types d'objectivité qui pourraient bien être à la base de la convergence de certains types d'évaluations.

Le tableau affectif met en valeur une objectivité de fait qui provient à la fois des niveaux de l'évaluation automatique et de l'activité réflexive ainsi que de l'interaction sociale. Pour ce qui est du premier niveau, les sentiments et plus particulièrement les réactions émotionnelles des êtres humains sont relativement bien coordonnés parce qu'ils résultent en partie de facteurs innés et parce que notre environnement social dirige nos réactions émotionnelles dans le sens de la consistance avec celles de nos voisins ; souvenons-nous par exemple des phénomènes de la contagion émotionnelle ou du besoin de maintenir des relations sociales plaisantes. De plus, ces sentiments et réactions émotionnels influencent largement nos évaluations réflexives ; la coordination se transmet ainsi du domaine de l'automatique au niveau réflexif. Comme nous l'avons vu, à ce niveau, des considérations d'ordre pratique ainsi que la demande de cohérence favorisent la convergence. Enfin, au sein de l'interaction sociale, il existe toute une batterie de mécanismes comme les biais du prestige ou du conformisme, qui assure une assez bonne cohérence de fait entre les valeurs et normes auxquelles souscrivent les sujets moraux.[266] Il vaut la peine de présenter brièvement ici une hypothèse ingénieuse

[266] Il est à noter que les données empiriques portent à croire que l'harmonie des valeurs sera atteinte moins grâce à des arguments logiques que par le moyen de la persuasion affective (voir HAIDT 2001, pp. 818-819 ; EDWARDS & VON HIPPEL 1995).

développée par Shaun NICHOLS : l'« hypothèse de la résonance affective ». Selon lui, les expériences émotionnellement chargées ont une plus grande probabilité de rester gravées dans la mémoire des individus[267] et transmises à d'autres membres du groupe. Ainsi, par exemple, une prohibition normative a plus de chances d'être stable au niveau de l'évolution culturelle si elle interdit des situations qui outrent généralement les gens (NICHOLS 2004, chap. 6). Ce mécanisme de sélection culturelle favoriserait également la cohérence entre les valeurs et normes auxquelles nous souscrivons.

En plus de cette objectivité *de fait* (même si elle est plus ou moins limitée), je pense que l'on peut parler d'objectivité *psychologique* ; cette forme d'objectivité a moins à voir avec la connaissance qu'avec le simple fait de considérer certaines valeurs, normes ou jugements comme intersubjectivement valables. Il est possible que ce phénomène puisse s'expliquer par un mécanisme psychologique renforcé par la prise de conscience d'un accord interpersonnel. Je m'explique. Les êtres humains sont dotés d'un mécanisme psychologique particulier, l'« empathie égocentrique », qui pourrait bien être à l'origine de leur croyance en l'objectivité des valeurs auxquelles ils souscrivent.[268] Le principe en est le suivant. Lorsque nous observons une action d'autrui, nous faisons l'expérience de cette action comme si nous en étions nous-mêmes l'auteur ; ce faisant, nous tenons uniquement compte de nos propres dispositions sans considérer l'état subjectif dans lequel se trouve la personne qui agit réellement.[269] Par exemple, les parents ressentent du dégoût lorsqu'ils surprennent leur enfant en train de manger ses propres excréments, même si l'enfant ne présente aucun signe de révulsion. Ce qui est intéressant avec l'empathie égocentrique, c'est qu'elle induit en nous la pensée que la manière dont nous évaluons les choses devrait être partagée de manière intersubjective. Et si par ailleurs, nous pouvons constater que les valeurs que nos sentiments nous incitent à adopter sont conformes aux comportements et valeurs exprimées par un grand nombre de nos voisins, alors nous disposons de bonnes raisons

[267] NICHOLS identifie différents mécanismes qui pourraient être responsables de l'impact des émotions sur la mémoire ; il mentionne notamment le fait que les stimuli émotionnels induisent une excitation qui cause probablement une augmentation de la production de glucose dont on sait qu'elle a pour effet de renforcer la mémoire (NICHOLS 2004, pp. 126-127).

[268] Pour une explication des origines évolutionnaires de l'empathie égocentrique, voir FESSLER et NAVARRETE 2003, pp. 15-16.

[269] Cette forme d'empathie est *égocentrique* parce qu'il n'est pas question de comprendre ce que ressent l'autre (p. 166), mais plutôt de ressentir ce que nous ressentirions si nous étions à sa place.

de considérer ces valeurs comme objectivement valables ; l'objectivité ressentie sous l'influence de l'empathie égocentrique est renforcée par la prise de conscience d'un accord intersubjectif.[270]

Il est intéressant de noter ici qu'au contraire des jugements de valeur spontanés (verbalisés ou non), les jugements sophistiqués et les normes sont justifiés, du moins dans l'esprit des gens qui les expriment ; ils reposent sur des valeurs et des normes considérées comme objectives.[271] Ainsi les conditions sont réunies pour que puissent s'engager les discussions interpersonnelles au sujet de ce qu'il est bien ou mal de faire.

5.2.2. Le tableau affectif de la motivation

Un problème important en philosophie morale et plus généralement en philosophie de l'action est de savoir ce qui nous motive à faire des choix et à agir en fonction de ces choix. Ainsi le tableau affectif des valeurs doit être complété par un tableau affectif de la motivation à l'action.

La question qui va m'intéresser dans cette section est celle de savoir si l'on peut présupposer que le simple fait d'adhérer à une norme, une valeur consciente[272] ou de produire un jugement de valeur est accompagné d'une tendance (pas forcément consciente) à agir en conformité avec cette norme, cette valeur consciente ou ce jugement. Si l'on répond par l'affirmative,[273] alors on sera dérangé par des situations

[270] Un peu dans la même ligne mais sans recourir à l'empathie égocentrique, Michael RUSE (1993/1991, p. 60) défend l'idée que certains sentiments (il les appelle « sentiments moraux ») sont tellement prégnants qu'ils engendrent en nous la croyance que les comportements vers lesquels ils nous poussent possèdent une valeur intersubjective. Par exemple, nous sommes naturellement enclins à réprouver les rapports incestueux et cette réprobation s'impose à nous sous forme d'un devoir (éviter et empêcher les rapports incestueux) qui incombe à la fois à nous et à autrui. Toutefois, l'analyse de RUSE est un peu fourvoyante, car d'une part il confond intersubjectivité et universalité, et d'autre part il lie l'objectivité psychologique à la croyance en l'existence d'une réalité morale externe (réalisme métaéthique). Or ce sont deux croyances distinctes qui ne s'impliquent pas forcément l'une l'autre (à ce propos, voir p. 328).

[271] Notons qu'en affirmant que les valeurs sont considérées comme *objectives*, je ne dis pas qu'elles sont considérées comme *universelles*. Le degré d'impression d'objectivité peut varier d'un cas à l'autre en fonction de l'implication émotionnelle du sujet et du degré d'accord intersubjectif dans son entourage.

[272] Il faut distinguer ici les valeurs intuitives qui guident nos actions et nos jugements spontanés, des normes et valeurs conscientes qui sont un produit de la réflexion des sujets.

[273] C'est la position défendue par exemple par Allan GIBBARD (2002/1990).

comme celle-ci. Imaginons que Reymond souscrit à la norme qui prescrit d'aider toute personne en détresse. Un jour, en rentrant chez lui, il voit un homme étendu de tout son long sur le trottoir. Il est clair que l'homme a été battu et qu'il souffre horriblement. Reymond n'est toutefois pas ému par la scène, peut-être parce qu'il reconnaît cet homme comme le clochard de son quartier, ou peut-être même sans raison particulière. Il décide de passer son chemin sans lui prêter garde, sachant parfaitement qu'en faisant cela, il transgresse une norme à laquelle il souscrit et qu'au fond il aimerait bien être capable de suivre. En d'autres termes, il agit consciemment à l'encontre de ce qu'il pense devoir faire.

Nous nous trouvons ici confronté à un cas parfaitement crédible mais assez troublant ; en effet, on aimerait pouvoir dire que le fait de souscrire à une norme et aux jugements qui en découlent est un facteur motivant à l'action. Dans cette section, je vais tenter de montrer qu'il faut accepter cette réalité ; les normes auxquelles nous souscrivons et les jugements qui en découlent n'ont aucun pouvoir motivant. Toutefois, comme nous le verrons, affirmer cela ne nous mènera pas à défendre la conclusion que l'évaluation et la normativité n'a rien à voir avec la motivation à l'action. Si les normes acceptées consciemment et les jugements moraux plus sophistiqués qui en dérivent ne sont pas motivants en eux-mêmes, je soutiendrai en revanche que la motivation est intrinsèquement liée aux jugements de valeur spontanés. Pour utiliser un jargon philosophique, la position qui sera défendue peut être qualifiée de « partiellement internaliste »[274], ou internaliste dans un sens restreint à certains états d'esprit : les jugements de valeur spontanés. Je reviendrai plus loin sur ces idées mais pour commencer, il vaut la peine de passer en revue un certain nombre de résultats empiriques sur lesquels repose le tableau affectif de la motivation.

i. Données empiriques relatives à nos choix moraux

Au cours de la dernière décennie, on a pu constater pour la question de la motivation, un gain d'intérêt significatif en psychologie, neurosciences et en économie

[274] Dans ce papier, les termes « internalisme » et « externalisme » seront utilisés dans un sens peu sophistiqué. Il s'agit simplement de la question de savoir si le fait de former un jugement motive à agir conformément à ce jugement ; l'internaliste répondra par l'affirmative alors que l'externaliste répondra par la négative.

expérimentale. Les résultats obtenus dans le cadre de ces sciences suggèrent fortement que les émotions sont un ingrédient essentiel pour prendre des décisions pratiques, c'est-à-dire pour motiver à l'action.

En neuroscience et économie expérimentale, des modèles tirés de la théorie des jeux ont été réalisés expérimentalement avec des sujets humains. Dans une étude utilisant le jeu de l'ultimatum, Alan SANFEY, Jim RILLING et collègues (2003) ont cherché à déceler les substrats neuronaux des processus cognitifs et émotionnels impliqués dans des situations de partage égal. Le jeu de l'ultimatum se pratique à deux joueurs. Les expérimentateurs donnent au premier joueur une certaine somme d'argent et lui demandent d'en donner une partie (pas forcément égale) au deuxième joueur. Le deuxième joueur peut accepter ou refuser la part qui lui est proposée (il connaît donc la somme). S'il accepte, la somme est divisée selon l'offre du premier joueur. S'il refuse, la somme totale est retirée du jeu si bien qu'aucun joueur n'obtient un gain ; dans ce cas, on peut dire du second joueur qu'il adopte un comportement punitif. Dans l'expérience de SANFEY et RILLING, les cerveaux des joueurs sont scannés pendant le jeu à l'aide de la technique d'imagerie cérébrale (les sujets ont été placés dans des tomographes à résonance magnétique). Les résultats de l'étude montrent que, comparées aux offres égales, les offres inégales ont pour effet, sur le deuxième joueur, d'augmenter sensiblement l'activité de l'insula antérieure, une région du cerveau associée aux émotions négatives. De plus les expérimentateurs observent une corrélation entre la force de la réponse émotionnelle négative et le rejet des offres inégales (c'est-à-dire le choix d'adopter un comportement punitif). Ces résultats suggèrent l'existence d'une étroite relation entre les épisodes émotionnels et la prise de décision. Ils trouvent écho dans une autre expérience menée par Ernst FEHR et Simon GÄCHTER (2002) sur la base d'un jeu similaire (le jeu du bien public). Dans le suivi de leur expérience, au moyen d'un questionnaire, ils ont demandé aux sujets ce qui les motive à punir de manière altruiste, c'est-à-dire à leurs propres frais et sans attente de compensation ultérieure. Sans grande surprise, les sujets répondaient systématiquement qu'il s'agissait de la colère éprouvée envers les opportunistes.

Les expériences menées dans le cadre des neurosciences et de l'économie expérimentale parviennent à montrer une corrélation entre les réactions émotionnelles et le comportement. Quoique l'hypothèse soit très plausible, il n'est toutefois pas certain que les réactions émotionnelles soient un ingrédient motivationnel *essentiel* ; elles pourraient simplement être corrélées. C'est dans le champ de la psychologie que l'on

trouve des raisons supplémentaires d'accorder une importance particulière aux émotions. Récemment, dans une étude basée sur la comparaison de personnes en bonne santé et de patients souffrant de troubles de régulation du comportement social, Jennifer BEER et collègues (2003) ont montré que la déficience du comportement social chez ces patients est associée avec un fonctionnement anormal des « *self-conscious emotions* », c'est-à-dire d'émotions liées à la conscience de soi comme la honte, l'embarras ou la fierté. Il semblerait donc que le fait de ressentir de manière appropriée les émotions liées à la conscience de soi est un facteur nécessaire au comportement social.

Dans la même veine, différentes études sur des psychopathes ou patients avec des lésions de la partie préfrontale du cerveau (DAMASIO 2001/1994, BLAIR *et al.* 2005, BECHARA *et al.* 1996, GRAY *et al.* 2003) révèlent que ces patients présentent un dysfonctionnement de leur système affectif sans pour autant souffrir de déficiences au niveau du raisonnement pratique. D'autre part, alors même qu'ils sont parfaitement capables de penser en termes de normes, ces patients ne montrent aucun intérêt pour les normes sociales et morales et se comportent de manière antisociale (à un degré nettement plus dramatique dans le cas des psychopathes). A nouveau, il semblerait que le bon fonctionnement des processus affectifs soit crucial pour l'action morale.

Récemment, Matteo MAMELI (2005) a analysé de manière détaillée les résultats des expériences de DAMASIO sur les patients avec lésions préfrontales. Il constate qu'en plus d'être incapables d'agir de manière sociale, ces patients ont également beaucoup de peine à prendre des décisions pratiques de tous les jours malgré le fait qu'ils soient tout à fait capables de mener des raisonnements logiques. Il en déduit que nos capacités de prendre des décisions pratiques en général, y compris les décisions morales, dépendent de manière cruciale du bon fonctionnement de notre système affectif ; en d'autres termes, nos choix pratiques sont le produit de nos émotions.

Toutes ces études empiriques suggèrent que l'activation de certaines émotions est une condition essentielle au comportement moral et plus généralement à la concrétisation des choix pratiques.[275] Nous disposons donc de bonnes raisons de penser qu'un facteur émotionnel doit être présent dans un jugement si l'on veut être motivé à agir conformément à ce jugement.

[275] Pour une bonne revue des recherches qui montrent que l'efficacité des normes sociales dépend d'émotions suffisamment fortes pour motiver à l'action, voir GREENE 2005.

ii. Une solution internaliste modérée

Les recherches empiriques laissent penser que les émotions sont essentielles aux prises de décisions pratiques, c'est-à-dire qu'elles sont un ingrédient motivationnel nécessaire. Deux conséquences en découlent. Premièrement les jugements de valeur spontanés sont intrinsèquement motivants puisque ce sont des réactions émotionnelles. Deuxièmement les produits du processus réflexif (les valeurs conscientes et normes acceptées ainsi que les jugements sophistiqués) ne sont pas motivants en eux-mêmes puisqu'ils ne possèdent aucune dimension émotionnelle. La motivation doit provenir d'une source externe ; ou bien d'une réaction émotionnelle corrélée au jugement sophistiqué, ou bien d'un désir d'obéir au jugement en question (par exemple parce que nous voulons éviter la punition ou soigner nos contacts).

Ainsi, à l'exemple de Reymond, il est possible de juger qu'une chose devrait être faite sans pour autant être motivé à le faire. Mais une telle situation ne peut surgir qu'en cas d'élaboration froide d'un jugement sophistiqué qui ne correspond pas à l'évaluation émotionnelle de la situation. Il est possible que Reymond n'aie eu aucune réaction émotionnelle particulière, ou s'il en a eu une correspondante à son jugement sophistiqué, alors elle a été supplantée par autre émotion contradictoire (par exemple la crainte de devoir ramener le clochard chez lui et de s'en occuper).

En philosophie morale, cette position hybride ou partiellement internaliste peut s'avérer assez dérangeante puisque précisément les jugements moraux paradigmatiques (ceux qui peuvent être fondés sur des normes et des valeurs) ne motivent pas à l'action. Dès lors, on pourrait penser que lorsqu'on agit de manière moralement vertueuse, c'est toujours par hasard. Il y a au moins trois raisons de penser que cette conclusion est trop pessimiste.

D'une part, nous avons vu que nos sentiments et réactions émotionnelles influencent largement nos évaluations réflexives. Ainsi, de fait, la plupart de nos normes, valeurs conscientes et jugements sophistiqués sont intiment corrélés à nos réactions émotionnelles. C'est d'ailleurs la raison pour laquelle la motivation altruiste (altruisme psychologique motivationnel) est si souvent confondue avec les intentions ou désirs altruistes conçues par les gens (altruisme psychologique sophistiqué).

D'autre part, les liens entre nos réactions automatiques et notre activité réflexive sont plus intimes que l'on aurait pu penser. Il y a non seulement l'influence des premières sur la seconde mais également l'effet inverse : les produits du processus

réflexif peuvent exercer un impact indirect sur nos réactions émotionnelles. Par exemple, en affinant notre compréhension cognitive de la situation jugée, nous pouvons déceler de nouvelles caractéristiques pertinentes qui vont modifier nos réactions émotionnelles ou en déclencher de nouvelles. De manière encore plus intéressante, je pense qu'il faut prendre au sérieux la possibilité de manipuler les valeurs ancrées dans nos esprits, celles qui causent nos réactions émotionnelles (à l'exemple de l'expérience du dégoût hypnotique mentionnée plus haut). Par exemple, si nous voulons qu'une norme ou une valeur consciente que nous acceptons ait une force motivante, il faut tenter de l'internaliser, de la graver dans notre esprit, de manière à ce qu'elle occasionne les réactions émotionnelles correspondantes (en faire une valeur intuitive) ; il s'agit donc de manipuler nos émotions de manière indirecte. Nous le faisons probablement déjà dans une certaine mesure sans même avoir à nous fixer consciemment ce but. Ces questions sont très complexes et délicates, si bien que je ne pourrai malheureusement pas les développer plus avant dans le cadre de ce travail.

Enfin, souvenons-nous de l'empathie égocentrique qui est à la base de l'objectivité psychologique. Un aspect central de ce mécanisme psychologique est le fait qu'il implique les émotions (les parents sont dégoûtés à la vue de leur enfant qui consomme ses propres excréments). Il s'ensuit que si l'objectivisme psychologique est effectivement lié à l'empathie égocentrique, alors les valeurs, les normes ou les jugements que nous considérons comme objectifs (au sens de l'objectivité psychologique) sont soutenus par nos émotions. On peut donc dire qu'en dépit d'une relation causale directe, ils sont indirectement motivants ou étroitement liés à une motivation. Il est important de noter ici que la motivation ne vient pas du fait de penser qu'une valeur, une norme ou un jugement est objectif ; la motivation provient du sentiment impliqué dans le mécanisme qui nous fait penser que cette valeur, cette norme ou ce jugement est objectif.

Bilan

Pour récapituler, le tableau affectif de l'évaluation montre que nous produisons différents types de jugements dans différentes circonstances. Les jugements de valeur spontanés sont des réactions émotionnelles, des manières de voir le monde qui, au fil de la vie, sont influencées par différents facteurs comme des tendances génétiquement

déterminées (parmi elles il y a des mécanismes émotionnels comme les émotions altruistes), les expériences individuelles passées ou le contexte socioculturel. Lorsque nous verbalisons nos jugements de valeur spontanés nous produisons de nouveaux jugements, cognitivement plus complexes, qui appellent à la justification. Lorsque nous justifions ces derniers sur la base de normes et de valeurs conscientes, nous produisons des jugements sophistiqués (jugements normatifs ou jugements de valeur). Ceux-ci, lorsqu'ils sont renforcés par l'empathie égocentrique, prétendent à une forme d'objectivité : l'objectivité psychologique (c'est d'ailleurs ainsi qu'ils peuvent être liés de manière indirecte à la motivation).

Quant aux valeurs conscientes et aux normes auxquelles nous souscrivons, le tableau affectif les présente comme des productions *post hoc* qui découlent de l'activité de rationalisation ou de justification de nos jugements spontanés. Et si elles sont internalisées, c'est-à-dire si elles sont gravées dans nos esprits de manière à diriger nos réactions émotionnelles, alors elles deviennent des valeurs intuitives qui doivent être comprises comme des mécanismes ou des raccourcis mentaux.

Enfin, le tableau affectif de la motivation montre que les normes auxquelles nous souscrivons et les jugements sophistiqués que nous exprimons n'exercent pas d'effet particulier sur nos actions si nous n'avons pas par ailleurs une attitude émotionnelle correspondante. La motivation est d'abord liée à une manière primitive et non réflexive d'interpréter le monde : le jugement de valeur spontané. D'autre part, j'ai défendu l'idée qu'il n'y a habituellement pas divergence entre les jugements de valeur spontanés et notre motivation à agir ; cela est simplement dû au fait que les jugements de valeur spontanés sont des réactions émotionnelles. Cette analyse a l'avantage de rendre compte du fait que les normes auxquelles nous souscrivons et les jugements sophistiqués qui leurs sont associés ne prescrivent pas toujours ce que nous choisissons en fin de compte de faire ; ils échouent à nous motiver lorsqu'ils ne sont pas soutenus par les jugements de valeur spontanés.

Notons pour finir que même s'il s'inspire de travaux sur les jugements et émotions considérés comme moraux, le tableau affectif décrit l'activité évaluative et normative dans son ensemble mais ne permet pas encore de distinguer, parmi l'ensemble des normes, valeurs et jugements que l'on peut produire, lesquels sont moraux et lesquels sont plus généralement d'ordre social, voire même d'intérêt personnel. Toutefois, étant donné l'importance capitale accordée aux réactions

émotionnelles, il faut prendre au sérieux l'hypothèse selon laquelle on peut trouver dans les émotions le moyen d'identifier le phénomène moral. Cette question sera traitée dans la section suivante.

5.3. Une analyse des émotions permet-elle de délimiter le champ de la moralité ?

La section précédente se conclut par la constatation que le tableau affectif ne fournit pas les clés nécessaires pour distinguer le champ de la morale de celui de la simple évaluation. Il faut trouver une autre voie. Etant donné l'importance accordée aux réactions émotionnelles dans l'activité évaluative, il vaut la peine de se demander s'il est possible d'y recourir pour définir le champ de la moralité. Peut-on mettre en évidence un critère qui permette de distinguer, parmi les réactions émotionnelles, lesquelles sont morales ? Etant donné qu'une réaction émotionnelle est une occurrence d'émotion, cette question est étroitement liée à celle de savoir s'il existe des émotions morales, et le cas échéant, comment comprendre la notion d'émotion morale. Dans cette section, je commencerai par présenter différentes compréhensions de cette notion. Aucune de ces lectures ne me paraissant convaincante, je poursuivrai sur une voie à première vue prometteuse qui consiste à tenter de trouver une explication fonctionnaliste de l'évolution des émotions morales ; il se pourrait que les émotions morales aient pour fonction évolutionnaire de soutenir les relations sociales coopératives. Cette explication sera testée en fonction de sa capacité d'établir une liste concrète d'émotions qui peuvent prétendre au qualificatif de « moral ». Après une élaboration de cette liste, il apparaîtra que l'approche fonctionnaliste évolutionnaire s'avère peu concluante. Je proposerai de sortir de cette impasse en suggérant que la moralité n'est qu'un label que l'on attribue aux émotions (ou plus justement aux épisodes émotionnels) lorsqu'elles apparaissent dans des contextes moralement pertinents. De cette discussion, il ressortira qu'il n'est pas possible de partir d'une analyse des émotions pour délimiter le champ de la moralité.

5.3.1. Ce que l'on peut entendre par émotion morale

Que faut-il entendre par émotion morale ? Il y a bien des manières de comprendre cette notion. En voici quelques-unes.

Certains auteurs assurent qu'il y a des émotions morales dans leur essence propre par opposition à des émotions immorales (à ce propos, voir DUMOUCHEL 2004 ; DEONNA 2007). Comme émotion morale on aurait par exemple l'amour du prochain, la culpabilité ou la compassion. Comme émotion immorale, on pourrait compter l'amour propre qui a été fustigé par ROUSSEAU (1999/1762), la jalousie ou le ressentiment dénigrés par NIETZSCHE (2000/1887) ou la vanité rejetée par HOBBES (2000/1651). Puisque ces émotions sont morales ou immorales en elles-mêmes, l'expérience des premières serait, par définition, toujours morale et l'expérience des secondes toujours immorale.[276]

Une autre lecture considère l'émotion morale comme un moyen heuristique de détecter la moralité de situations ou comportements. Selon cette approche, on attribuerait une fonction aux émotions morales : celle de révéler des valeurs (HUTCHESON 1991/1726, p. 67 ;[277] TAPPOLET 2000[278]) ou de révéler la conformité d'une situation à des normes morales. Dans ce contexte, les émotions sont considérées comme morales précisément parce qu'elles sont un outil cognitif indispensable à l'activité morale.

Une troisième lecture consiste à faire des émotions morales un critère d'individuation de la moralité. C'est l'option choisie par certains philosophes. Pour

[276] Notons cependant que cette caractérisation est sujette à d'innombrables contre-exemples. Par exemple, il n'est pas si évident d'affirmer qu'il est moral d'éprouver de l'amour pour Hitler. Inversement, pourquoi ne pourrait-on pas admettre qu'il existe des formes de jalousie neutres du point de vue moral ? On trouvera même des auteurs qui affirment que la jalousie est une émotion morale paradigmatique dans la mesure où elle est ressentie dans les bonnes circonstances (KRISTJANSSON 2002).

[277] Pour Francis HUTCHESON (1991/1726) par exemple, nous possédons un sens moral qui s'exprime par le biais d'émotions ; plus précisément, il existerait deux types généraux d'émotions morales, l'approbation et la condamnation, qui nous permettraient de saisir le bien et le mal.

[278] Selon Christine TAPPOLET, les émotions sont comparables aux expériences perceptuelles ; elles révèlent une réalité axiologique et possèdent des conditions d'adéquation. Par exemple, une émotion d'admiration sera adéquate si et seulement si cette personne est réellement admirable. Des thèses similaires sont défendues par Ronald DE SOUSA (1987), Peter GOLDIE (2000), Sabine DÖRING (2007).

Alfred AYER (1946/1936) par exemple, l'évaluation morale consiste en l'expression d'une émotion d'approbation ou de désapprobation ; selon Allan GIBBARD (2002/1990), les jugements moraux se caractérisent par le fait qu'ils portent sur les émotions morales de colère et de culpabilité. Ici à nouveau, ces émotions sont considérées comme morales dans leur essence propre mais à la différence de la première lecture, il ne s'agit pas d'une moralité justifiée.

Cette énumération des manières de concevoir les émotions morales n'est pas exhaustive. Je mentionnerai plus loin d'autres positions qui me paraissent plus prometteuses (notamment l'idée que les émotions peuvent uniquement être considérées comme morales de manière extrinsèque). Pour le moment, voyons ce qu'il en est de celles-ci.

Comme le fait bien remarquer Paul DUMOUCHEL (2004), ce foisonnement d'interprétations des émotions morales est assez gênant, d'autant plus qu'aucune de ces lectures ne permet de décider quelles émotions ont droit au qualificatif de « moral ». Pour décider de la meilleure lecture, il serait intéressant de disposer d'arguments supplémentaires et ne pas se contenter d'affirmations arbitraires. Il faudrait quelque chose de plus que la seule cohérence du système philosophique érigé autour d'une interprétation particulière de l'émotion morale.

Une manière peut-être plus prometteuse d'aborder la notion d'émotion morale serait d'adopter une perspective fonctionnaliste évolutionnaire. En effet, pour peu que l'on veuille faire de la morale un phénomène significatif à l'échelle humaine, toutes les lectures proposées ci-dessus de la notion d'émotion morale doivent admettre que les êtres humains en général possèdent ces émotions morales (du moins dans des circonstances normales), c'est-à-dire que ces émotions sont universelles à l'échelle humaine. D'autre part, lorsqu'il s'agit d'expliquer leur présence universelle à l'échelle humaine, il n'y a guère d'autre possibilité que de les considérer comme innées[279] ou au moins de postuler des capacités innées qui en permettent le développement avec une très haute probabilité. Si on ajoute à cela une perspective évolutionnaire, cela nous mène directement à l'idée que les émotions morales sont des produits de l'évolution naturelle, probablement sélectionnées parce qu'elles produisent un avantage sélectif.

[279] Je comprends inné au sens où il doit exister quelque facteur héréditaire qui régit la production des émotions morales chez les êtres humains.

Plus précisément, comme le fait remarquer Luc FAUCHER, « l'expression 'émotions morales' semble suggérer en effet qu'un groupe d'émotions a été sélectionné pour répondre aux problèmes du domaine moral » (2007, p. 113). On trouve ici non seulement l'idée que les émotions morales sont des entités suffisamment homogènes pour pouvoir être des objets de sélection, mais qu'en plus, elles ont toutes été sélectionnées pour la même raison ; en d'autres termes, les émotions morales pourraient être comprises comme une seule sorte d'objet de sélection. Dès lors, la tâche est de déterminer en vertu de quelles caractéristiques elles ont été sélectionnées, ou vu autrement, pour répondre à quels problèmes de notre environnement adaptatif elles ont évolué. Si l'on parvient à expliquer cela, on disposera du même coup d'un bon critère pour définir les émotions morales, et comme on le verra, ce critère pourrait bien s'accommoder d'une lecture proposée ci-dessus plutôt que d'une autre.

Dans ce qui suit, je vais présenter l'explication évolutionnaire classique des émotions morales selon laquelle elles auraient pour fonction de soutenir les relations sociales coopératives. Je chercherai ensuite à déterminer quelles sont les émotions qui, selon une perspective fonctionnaliste évolutionnaire, peuvent prétendre au qualificatif de moral. Après une élaboration grossière de cette liste, il apparaîtra que l'explication classique de l'évolution des émotions morales s'avère insuffisante. Je proposerai de sortir de cette impasse en développant une « solution minimaliste ».

5.3.2. La fonction coordinatrice et coopérative des émotions morales

Avant de présenter l'hypothèse évolutionnaire classique relative à ce qui confère la moralité aux émotions, il me paraît important de dire quelques mots sur le phénomène émotionnel en général. Ce dernier est généralement conçu selon un modèle tripartite : certaines causes typiques vont déclencher certaines réactions physiologiques et cognitives typiques lesquelles sont liées à des tendances typiques à l'action. Par exemple un épisode de colère est une réaction à un ensemble de circonstances comme le fait de subir des injures ou recevoir des insultes. Cet épisode de colère se caractérise par un certain état du système neuronal et endocrinien, ainsi que par une expression faciale spécifique à la colère. Enfin, un sujet en colère sera motivé à accomplir des actes punitifs.

Selon les théoriciens évolutionnistes, ces trois états font partie d'un plan génétique propre à toute émotion. Quant aux détails du plan (c'est-à-dire, pour une émotion particulière, l'ensemble de ses causes, son type d'expression et de tendances à l'action), il sera en partie génétiquement déterminé et en partie modulé par les influences de l'environnement et par les expériences personnelles des sujets. Certaines émotions plus basiques comme les formes primitives de colère ou de peur sont profondément ancrées dans nos gènes alors que des émotions plus sophistiquées comme la honte ou la culpabilité sont l'effet des influences culturelles (l'environnement). Cette explication rend compte du fait que certaines émotions se rencontrent dans n'importe quelle culture humaine[280] alors que d'autres s'expriment de manière plus locale.[281]

Certains auteurs (GIBBARD 2002/1990 ; FESSLER & HALEY 2003 ; TRIVERS 1985) ont suggéré que les émotions morales ont évolué parce qu'elles permettaient de renforcer la coopération et la coordination entre les personnes d'un groupe social. En d'autres termes, elles ont été sélectionnées pour leurs effets bénéfiques sur la vie sociale et leur fonction biologique consiste à coordonner les actions et les attentes des gens sur un modèle coopératif.[282] D'autre part, ces auteurs pensent souvent en termes d'émotions complémentaires. Par exemple, l'émotion complémentaire de la colère morale[283] pourrait être la culpabilité. Elles sont complémentaires parce que les actions punitives que la colère nous pousse à accomplir (il s'agit des expressions typiques de la colère) concordent avec les punitions que l'on est prêt à supporter lorsque l'on se sent

[280] Par exemple les émotions basiques définies par Paul EKMAN et Wallace FRIESEN (1989): colère, dégoût, peur, joie, tristesse ou surprise.

[281] Par exemple, dans les cultures japonaises on trouve l'émotion « amae » qui consiste à se sentir agréablement dépendant de l'amour d'autrui, ou dans l'île du Pacifique Ifaluk, les gens font l'expérience de l'émotion « fago », un mélange de compassion, d'amour et de tristesse (voir ROBINSON 2004, p.39).

[282] Notons que cette thèse, associée à l'idée que les émotions morales jouent le rôle de critère d'identification de la moralité, mène à penser que la fonction de la moralité elle-même est de coordonner les actions humaines et favoriser la coopération (cette position est notamment défendue par GIBBARD 2002/1990). Pour une critique de ce genre d'approches, voir section 5.1.

[283] Suivant Allan GIBBARD (2002/1990), considérons la colère morale comme une catégorie assez large incluant des émotions comme le ressentiment ou l'indignation.

coupable. Ainsi si ces deux émotions sont répandues dans une population, elles favoriseront grandement les interactions saines et coopératives.[284]

Robert TRIVERS par exemple pense que la culpabilité implique un jugement selon lequel il est justifié que nous soyons punis ; plus précisément cette émotion sert à inhiber nos mécanismes défensifs, à accepter la punition et à favoriser les expressions d'excuses, tout cela ayant évidemment pour effet de rétablir les bons rapports avec la personne en colère (1985, p. 389). Dans la même ligne, Dan FESSLER et Kevin HALEY (2003) défendent l'idée que la culpabilité est causée lorsque l'on comprend qu'un de nos comportements pourrait causer du tort à une relation de coopération mutuelle. Cette émotion nous pousserait ensuite à des gestes de compensation pour le tort causé afin de réinstaurer un rapport de confiance.

Cette approche évolutionnaire fonctionnaliste nous fournit un critère simple (la coordination des interactions sociales coopératives) qui devrait permettre de définir une liste des principales émotions morales.

5.3.3. Une liste des principales émotions morales

Parmi les théoriciens qui adoptent une perspective fonctionnaliste et évolutionnaire des émotions morales, la question de savoir quelles émotions entrent dans la catégorie morale fait l'objet de débats.

Certains auteurs considèrent que seules les émotions négatives[285] peuvent vraiment prétendre entrer dans cette catégorie. C'est la position défendue par Allan

[284] « Take guilt and resentment: if one person resents an action of another and the other does not feel the corresponding guilt, we may expect trouble. Guilt makes possible the acknowledgment of wrong, and such modes of reconciliation as restitution, compensation, apology, and forgiveness. One's chances of damaging conflicts are reduced, then, if one feels guilty when guilt and its normal accompaniments are demanded by others, and if one demands guilt and its normal accompaniments only when others are prepared to feel guilty. Hence it tends to be advantageous for an individual to coordinate his guilt with the resentment of others and his resentment with the guilt of others. » (GIBBARD 2002/1990, pp. 67-68)

[285] Il y a plusieurs manières de définir les émotions positives et négatives. Je suivrai ici Aaron BEN-ZE'EV (2000, p. 72) pour qui les émotions positives incluent une évaluation positive, un sentiment plaisant ainsi que le désir de maintenir la situation qui en est la cause (et vice versa pour les émotions négatives). Cela implique de considérer comme négatives des émotions comme la sympathie ou la compassion.

GIBBARD. Pour lui, la moralité, dans son sens restreint, est conceptuellement liée aux émotions de culpabilité et de colère et non à d'autres émotions. Il reconnaît bien la pertinence morale d'une série d'autres émotions comme l'admiration, le respect, la pitié, la bienveillance (2002/1990, p. 291), mais selon lui, ces dernières joueraient plutôt un rôle de soutien et pourraient uniquement être considérées comme morales dans la mesure où elles permettent de soutenir l'action et affiner les deux émotions morales par excellence : la colère et la culpabilité.[286] La raison qui le pousse à accorder une telle place à la colère et à la culpabilité vient de ce que ces deux émotions sont corrélées. Comme nous l'avons vu plus haut, les causes de la colère sont les mêmes que celles de la culpabilité et les tendances à l'action de la culpabilité sont complémentaires de celles de la colère : la culpabilité calme la colère et cette alchimie permet de restaurer la coordination et la coopération. Selon GIBBARD, ce couple d'émotions est extrêmement puissant pour nous motiver à coopérer et ne pas tricher dans le cadre de relations sociales et coopératives. D'autre part, il est convaincu que ces deux émotions négatives sont nettement plus puissantes que les émotions positives comme la bienveillance ou le respect pour maintenir la coordination sociale. (2002/1990, pp. 257-269).[287] Voyons s'il est justifié à croire cela.

Le choix de la colère comme émotion morale par excellence semble assez justifié. Du moins, un bon nombre de données empiriques le confirment.

On peut montrer par exemple que la colère décourage la tricherie dans le contexte des relations sociales parce qu'elle est associée à une tendance à punir (étant entendu que la punition est une expression typique de la colère). C'est du moins dans ce sens que vont les résultats des recherches menées en anthropologie évolutionniste (Robert BOYD et collègues) et en économie expérimentale (Ernst FEHR et collègues) sur la

[286] Notons que cette conception de la moralité centrée sur deux émotions négatives est assez gênante car elle pousse GIBBARD à admettre que les individus de sociétés qui ne connaissent pas l'une de ces émotions (ou les deux) sont incapables de penser et agir moralement (du moins au sens étroit du terme). Or même si la colère morale semble se trouver dans toutes les cultures, il n'est pas certain que ce soit le cas pour la culpabilité (à ce propos, voir SCHERER & WALLBOTT 1994; FESSLER & HALEY 2003).

[287] Je pense toutefois qu'il y a une raison plus profonde à ce choix de restreindre le champ des émotions morales. Cela tient à la position que GIBBARD défend en philosophie morale : pour se démarquer des faiblesses philosophiques d'une autre position, l'émotivisme (AYER 1946/1936), GIBBARD fait de la colère et de la culpabilité un critère simple d'individuation de la morale. Je ne m'attarderai toutefois pas sur cette question ici (pour plus de détails voir CLAVIEN, soumis).

punition altruiste. Rappelons que la punition altruiste consiste à sacrifier une partie de ses ressources pour punir des comportements opportunistes sans attendre de bénéfices ultérieurs en retour (voir section 2.3.4). Du point de vue théorique, il est clair qu'en forçant les opportunistes à coopérer (sous menace de sanction), la punition altruiste engendre des effets bénéfiques pour la coopération et l'entraide au sein de grands groupes (BOYD *et al.* 2003 ; FEHR & FISCHBACHER 2003). Des expériences empiriques simulant des interactions sociales (où l'on demande à des sujets humains de pratiquer différents jeux comme le dilemme du prisonnier, le jeu de la confiance ou le jeu du bien commun) ont montré que s'ils ont la possibilité de punir les conduites opportunistes, les gens se comportent souvent comme des altruistes punisseurs, c'est-à-dire qu'ils sont prêts à punir les opportunistes à leurs propres frais et sans attente de bénéfices en retour (FEHR & GÄCHTER 2002 ; FEHR & FISCHBACHER 2004a). L'existence d'un lien étroit entre l'émotion de la colère et cette tendance à pratiquer la punition altruiste a également été montrée par FEHR et GÄCHTER. A la fin des sessions de jeux, ils ont distribué un questionnaire aux sujets leur demandant de s'imaginer, dans le cadre d'une situation de bien public, quelles émotions ils ressentiraient en tant qu'opportuniste ou en tant que punisseur. Les résultats montrent que la punition est motivée par la colère dirigée contre les opportunistes et que ces derniers anticipent les réactions de colère. Ces observations soutiennent l'idée que la colère est un facteur important sous-jacent à la punition altruiste, et par conséquent, qu'elle a un rôle crucial à jouer pour le maintien de groupes coopératifs.

Qu'en est-il de la culpabilité ? Au contraire de la colère, il y a peu de recherches évolutionnaires sur la culpabilité. Cela est probablement dû au fait que le concept de culpabilité est plus vague que celui de colère.[288] En consultant la littérature évolutionnaire, on se trouve constamment confronté à des définitions contradictoires de la culpabilité et des émotions proches comme le remords, le regret ou la honte. Cette confusion terminologique nous impose de considérer ces émotions comme une seule classe assez floue.[289] Mais cela n'empêche pas de se demander si ce type d'émotions

[288] Il n'est par exemple pas évident de faire la distinction entre les expressions faciales caractéristiques des épisodes de culpabilité et de honte.

[289] Evidemment, les anthropologues, psychologues ou philosophes ne seront pas très heureux de voir mettre dans le même panier des émotions comme la culpabilité ou la honte qui reposent manifestement sur des ressorts intellectuels différents (à ce propos, voir DEONNA 2007).

favorise la coopération et la coordination. FESSLER et HALEY (2003) pensent que la honte[290] induit une punition subjective en réaction à la transgression de normes et de ce fait, incite les individus à suivre les normes. Dans une société régie par des normes, une telle émotion a pour effet de réguler les comportements pro-sociaux.

Dans la même ligne, les économistes Samuel BOWLES et Herbert GINTIS (2002) ont mené des expériences de jeu de bien public afin d'explorer le rôle de la culpabilité et de la honte sur les comportements pro-sociaux. Dans le cadre de leurs expériences, ils ont focalisé leur attention sur les réponses données par les sujets qui ont été punis pour leur manque de volonté coopérative. Leur étude montre que le comportement des gens est influencé de manière significative par leurs sentiments de culpabilité, honte et regret. Eux aussi en déduisent que ce type d'émotions joue un rôle central dans les relations coopératives.[291]

Jusqu'ici, GIBBARD semble avoir raison. Toutefois, en science évolutionniste et psychologie sociale, on ne trouve pas uniquement des explications de la colère et des ses émotions complémentaires. Les théoriciens se réfèrent également à d'autres émotions négatives comme le dédain, le dégoût[292] ou des émotions positives comme l'amour, la gratitude, la fierté, l'admiration ou l'élévation (FESSLER & HALEY 2003; HAIDT 2003).

Récemment par exemple, dans une étude basée sur la comparaison entre des personnes en bonne santé et des patients souffrant de troubles de régulation du comportement social, Jennifer BEER et collègues (2003) ont montré que la déficience du comportement social chez ces patients est corrélée avec l'absence d'émotions comme la honte ou l'embarras. Comme second exemple, mentionnons le primatologue Frans DE WAAL. Selon lui, si l'on veut trouver les éléments clés de la moralité et comprendre

[290] Ils conçoivent la honte comme une émotion négative déclenchée par la prise de conscience qu'autrui sait que nous avons agi de manière moralement répréhensible.

[291] La clé de leur argumentation consiste à montrer que le comportement coopératif ne peut pas être expliqué par le seul fait de craindre la punition en cas de non-coopération.

[292] A première vue, il peut sembler étrange de compter le dégoût parmi les émotions morales. On pourrait se demander s'il ne vaudrait pas mieux parler d'indignation plutôt que de dégoût moral. Il semblerait que non car dans une expérience utilisant l'imagerie cérébrale par résonnance magnétique (J. MOLL *et al.* 2005), des chercheurs en neurosciences ont pu montrer que le dégoût ordinaire et le dégoût moral activent à peu près les mêmes zones du cerveau (à ce propos, voir aussi FAUCHER 2007).

comment la pensée et le comportement moral sont apparus au cours de l'évolution, il faut chercher des indices dans l'étude des espèces analogues à la nôtre comme les primates ou plus particulièrement les grands singes (1997/1996). Même si ces derniers ne peuvent pas être considérés comme des êtres moraux à part entière, ils font preuve d'un sens de la régulation sociale convergeant avec les valeurs intuitives qui régissent le comportement humain. DE WAAL montre également que la capacité de l'empathie cognitive (la faculté de comprendre les besoins et les émotions d'autres individus ; voir p. 166) et l'émotion de sympathie qui en résulte peuvent être observées chez les grands singes ; les chimpanzés par exemple s'engagent dans des activités de consolation. L'empathie et la sympathie semblent également essentiels pour le bon fonctionnement de leur vie sociale (c'est pourquoi DE WAAL les qualifie de proto-morales) ; elles promeuvent la cohésion, la coopération et les liens sociaux (FLACK & DE WAAL 2000).

De manière plus générale, Jonathan HAIDT (2003) distingue quatre familles d'émotions morales qui pourraient jouer un rôle fonctionnel significatif en matière de coordination et de coopération : la famille des émotions condamnatrices (dédain, colère, dégoût), la famille des émotions qui impliquent la conscience de soi (honte, embarras, culpabilité), la famille des émotions déclenchées par la souffrance d'autrui (compassion) et enfin la famille des émotions élogieuses envers autrui (gratitude, élévation).[293] Selon HAIDT, ces émotions sont déclenchées par des causes désintéressées et incitent à produire des actions pro-sociales. Parmi elles, les émotions les plus typiquement morales sont l'élévation, la compassion, la colère et la culpabilité. Le cas de l'élévation est particulièrement intéressant. Il semblerait qu'il s'agisse de l'émotion complémentaire du dégoût social ; c'est un sentiment positif et chaleureux ressenti lorsque nous sommes témoin d'actes que nous considérons comme moralement bons et qui nous incite à aider autrui et à nous en rapprocher (HAIDT 2000). HAIDT et collègues ont mené plusieurs expériences psychologiques sur l'émotion d'élévation. Les résultats montrent que lorsque les gens sont témoins d'actions moralement admirables, ils ressentent un fort désir d'être l'auteur d'actions similaires et de se rapprocher des gens qui se comportent ainsi.[294]

[293] Les termes utilisés par HAIDT pour ces quatre familles d'émotions sont: *other-condemning family*, *self-conscious family*, *other-suffering family* et *other-praising family*.

[294] Pour une explication fonctionnelle similaire relative à la pertinence morale de l'émotion de gratitude, voir MCCULLOUGH *et al*. (2001).

Notons qu'en comparaison avec les émotions négatives comme la colère, la tristesse ou la peur, les émotions positives ont été un peu négligées dans la littérature évolutionnaire. Selon Barbara FREDRICKSON (2000 ; 2003) qui a consacré plusieurs années à l'étude des émotions positives, la raison tient au fait qu'en comparaison des émotions négatives, il est plus difficile de les différencier les unes des autres, si bien que leur étude n'est pas aisée. La colère et la peur sont facilement reconnaissables parce que chacune est liée à une expression faciale paradigmatique ; en revanche les expressions faciales des émotions positives sont très similaires (elles s'expriment toutes par le sourire de Duchenne). De plus, les scientifiques étudient généralement les émotions en les associant à des tendances à l'action qui leurs sont propres ; par exemple, la colère incite à l'attaque ou à la punition alors que la peur pousse à la fuite. Ces tendances à l'action caractéristiques facilitent les scénarios explicatifs de l'évolution de ces émotions en termes d'adaptation. Malheureusement, il n'est pas aussi évident que les changements physiologiques et les tendances à l'action associées aux émotions positives soient favorables à la survie. Pour cette raison, FREDRICKSON propose d'abandonner ce cadre d'analyse étroit lorsqu'il s'agit d'étudier les émotions positives. Selon elle, les émotions positives ne résolvent aucun problème de survie immédiat. Par contre, elles exercent une influence profonde et durable sur la santé et le bien-être des gens, ainsi que sur leur capacité de gérer des situations difficiles, ce qui est également très favorable d'un point de vue évolutionnaire. Ainsi au niveau social, les émotions positives sont également un facteur de cohésion significatif. Concernant la question des émotions moralement pertinentes, FREDRICKSON pense que la fierté, la gratitude ou l'élévation incitent les gens à agir en faveur d'autrui. En créant une chaîne d'événements chargés d'une valence positive, elles peuvent déclencher et soutenir des spirales de comportements sociaux capables de faire d'une communauté une organisation sociale morale et harmonieuse (2003, p. 335)

En bref, une émotion positive comme l'élévation pourrait bien être un instrument aussi puissant que la colère ou la culpabilité pour maintenir la coopération et les comportements pro-sociaux. Dans un contexte social, elle est évolutionnairement avantageuse pour la communauté autant que pour les membres individuels de cette communauté puisqu'elle permet d'augmenter le taux d'assistance mutuelle ; et au niveau individuel, elle fournit énergie et bien-être.

Pour conclure, à la lumière des analyses évolutionnaires récentes, il semblerait que la liste des émotions morales (si l'on accepte le critère de la coopération et de la

coordination) soit assez large et inclue à la fois des émotions négatives et positives. Ainsi il est difficile de donner raison à GIBBARD lorsqu'il donne la priorité à la colère et à la culpabilité ; même si ces émotions semblent être des conditions nécessaires à la coopération, nous avons de fortes raisons de penser que la réduction de la moralité à elles deux et plus généralement au domaine du sanctionnable est par trop réductrice.

Il apparaît donc que la liste de émotions morales, même si elle ne semble pas close (certaines émotions pertinentes n'ont peut-être pas encore fait l'objet d'analyses évolutionnaires et il est probable que de nouvelles émotions morales surgissent au gré des variations culturelles et sociales), est très clairement limitée ; elle se restreint aux émotions qui favorisent la coordination et la coopération.

5.3.4. Les limites du critère de la coordination et de la coopération

Demandons-nous maintenant ce qu'il en est de l'idée selon laquelle les émotions sont morales seulement si elles ont un rôle significatif à jouer en matière de coopération et de coordination. Quoiqu'extrêmement tentante, j'aimerais montrer que cette réduction de la fonction des émotions morales à la coordination des interactions sociales coopératives devrait être remise en question.

Une première critique très générale et qui s'applique en réalité à toute théorie fondée sur une conception des émotions morales est de dire que si l'on jette un œil aux différentes positions morales qui ont été défendues au cours de l'histoire de la philosophie, force est de constater que les penseurs ne s'accordent pas sur la liste des émotions morales paradigmatiques. Cela semble suggérer que nul ne sait quelles sont les émotions morales (à ce propos, voir DUMOUCHEL 2004). Certes, cette objection reste assez superficielle : d'une part une absence de consensus n'implique pas qu'aucune option n'est dans le vrai ; d'autre part le théoricien évolutionniste pourrait répondre que ses explications fournissent précisément un moyen de décider lesquelles sont plus plausibles parmi les différentes positions existantes. Concédons-lui cela.

Une autre critique serait de dire que toutes les occurrences des émotions morales n'ont pas pour effet de produire de la coordination et de la coopération. On peut par exemple être en colère de manière complètement injustifiée et cela ne favorisera certainement pas les bons rapports sociaux. A cela, on pourrait répondre que les théories

évolutionnistes ne traitent pas des situations particulières mais plutôt des effets sur le long terme de l'occurrence régulière des émotions considérées. De plus, si une émotion a évolué pour remplir une fonction particulière (en l'occurrence renforcer la coopération, ce qui lui confère le label moral) il ne s'ensuit pas que toutes ses occurrences remplissent effectivement cette fonction.[295] Cela étant admis, il s'agit ensuite d'affiner le système moral pour définir quand l'occurrence d'une émotion morale est justifiée ou non ; il est important de distinguer ici entre le fait d'être moral et le fait d'être moralement justifié. L'intérêt de cette réponse est que l'approche évolutionnaire semble permettre ici de décider, parmi les différentes manières de concevoir les émotions morales listées ci-dessus, laquelle est la plus pertinente. Il semblerait que la lecture de l'émotion morale en tant que critère de moralité gagne du terrain ; selon cette ligne de pensée, un jugement devient *moral* en vertu du fait qu'il implique une émotion morale. Il faut ensuite recourir à d'autres critères pour décider, parmi les jugements moraux, lesquels sont justifiés (ou moralement corrects) et lesquels ne le sont pas ; c'est d'ailleurs ici que l'approche évolutionnaire doit être complétée par la philosophie traditionnelle.[296]

Pour remettre en question de manière convaincante le dogme de l'émotion morale comme fleuron de la coopération, il me semble que c'est dans le champ même des théories évolutionnistes que l'on peut trouver deux arguments aussi simples que percutants.

Tout d'abord, il n'est pas certain que toutes les émotions que les théoriciens évolutionnistes aiment à considérer comme morales favorisent la coordination et la coopération. C'est le cas par exemple du dégoût moral. Une instanciation caractéristique de cette émotion est la réaction face à l'inceste.[297] Or il n'est pas du tout certain que ce type de réactions peut inciter d'une quelconque façon à la coopération (voir FAUCHER 2007). L'hypothèse explicative la plus convaincante attribue au dégoût face à l'inceste des racines évolutionnaires différentes : on sait que les couples consanguins ont une

[295] En effet l'évolution est une affaire de probabilité ; un trait ne doit pas être adaptatif dans toutes ses occurrences pour être sélectionné, il suffit qu'il le soit assez souvent pour pouvoir être transmis avec une grande probabilité à la génération suivante.

[296] Pour une exemplification brillante de ce type d'analyse, voir GIBBARD 2002/1990.

[297] De manière sans doute assez ironique, c'est précisément Jonathan HAIDT (2001) qui a beaucoup travaillé sur cette émotion.

plus haute probabilité d'engendrer des enfants qui souffrent de maladies graves, si bien que sur le long terme, ce genre d'unions s'avère défavorable dans l'ordre biologique. Ce phénomène a favorisé l'émergence, dans beaucoup d'espèces, de dispositions rendant improbables les relations sexuelles entre proches parents. Chez certains animaux on peut observer le phénomène de dispersion des petits après le sevrage ; chez les êtres humains c'est le mécanisme de dégoût face à l'inceste qui a évolué. Ainsi, si le dégoût est une émotion morale, l'explication de l'évolution des émotions morales en termes de coordination et de coopération ne semble être au mieux que partielle.

Le deuxième argument susceptible de mettre en doute le dogme de l'émotion morale comme fleuron de la coopération prend le problème par l'autre bout. Certaines émotions semblent favoriser les interactions sociales alors même que de l'avis du sens commun, elles sont complètement étrangères à la moralité. Prenons l'exemple du sentiment de soumission : savoir se soumettre à plus fort que soi favorise clairement la coordination des relations sociales et toute une gamme de relations de coopération (même si cette coopération procède de manière inégale). Il en va de même d'autres émotions comme l'assurance de soi par exemple.

En fin de compte, il me semble que les difficultés liées à la recherche d'un critère évolutionnaire pour déterminer quelles sont les émotions morales relèvent du fait que la catégorie d'émotion morale comprise comme une seule sorte d'objet de sélection pose problème ; les émotions considérées comme morales n'ont pas toutes évolué parce qu'elles favorisent la coordination et la coopération mais pour des raisons qui restent propres à chacune d'entre elles.

5.3.5. En faveur d'une lecture minimale

Face aux insuffisances des théories présentées jusqu'à présent, je suggèrerai d'abandonner l'idée d'émotions intrinsèquement morales au profit d'une lecture minimaliste qui peut se décliner sous deux variantes dont je retiendrai la seconde.

La première variante de la lecture minimaliste consiste à dire que la moralité est une caractérisation que l'on applique à un ensemble d'émotions ; plus précisément, c'est

simplement à l'aide de nos outils conceptuels que nous classons une pluralité d'objets sous le label « émotion morale ». Il y a deux manières d'attribuer ce label.

Nous pouvons le faire en fonction d'un certain nombre de critères qui nous paraissent décisifs (sans pour autant qu'ils soient pertinents du point de vue évolutionnaire). Par exemple, une émotion pourrait être considérée comme morale si elle remplit la condition d'être systématiquement déclenchée par des situations qui (aux yeux des sujets) impliquent le bien-être et les besoins fondamentaux d'un ou plusieurs individus. [298]

Une autre manière serait d'attribuer le label moral aux émotions qui sont impliquées dans un contexte moralement pertinent. Par exemple la gratitude serait morale parce qu'elle est déclenchée lorsqu'on est témoin d'une action moralement bonne ; ou alors la sympathie serait morale parce qu'elle induit des comportements moraux.

Quoique plus convaincante, il me semble que cette lecture des émotions comme extrinsèquement morales a ses limites. En effet, vouloir attribuer la moralité à une émotion en tant que telle (c'est-à-dire à toutes ses occurrences) ouvre la voie aux contre-exemples ; on pourra alléguer par exemple que la sympathie envers des personnes exécrables comme HITLER ne peut pas être considérée comme morale.

Pour cette raison, je suggère d'adopter une seconde variante de la solution minimaliste qui revient à attribuer le qualificatif de moral, non aux émotions elles-mêmes mais uniquement aux occurrences d'émotions qui, ou bien remplissent les critères définis, ou bien interviennent dans des contextes que l'on considère comme moraux.

Même s'il ne parle par explicitement d'épisodes émotionnels, il me semble que la position de Luc FAUCHER illustre assez bien l'utilisation de critères externes pour classer les épisodes émotionnels sous l'enseigne de la moralité. Les critères qu'il invoque sont le fait d'être déclenché dans des contextes sociaux où les agissements des

[298] John RAWLS (1997/1971, p. 576) semble défendre une position de ce type. Selon lui, le ressentiment est une émotion morale parce qu'on l'éprouve toujours dans les situations où nous sommes victime d'une injustice.

uns ont des effets sur d'autres individus[299], d'être motivé par la conformité ou la non-conformité à une norme et enfin d'être désintéressé (2007, p. 114).

Notons que le choix de ce type de critères semble dépendre directement d'une conception préalablement fixée de ce qu'est la moralité. En conséquence, il ne me paraît pas forcément nécessaire de définir des critères d'individuation morale pour les épisodes émotionnels eux-mêmes. Je pencherais donc plutôt pour la deuxième manière d'attribuer le label de la moralité aux émotions : elles deviennent morales en quelque sorte par effet de répercussion, lorsqu'elles sont impliquées dans des contextes où des questions de nature morale sont posées. Par exemple, un épisode émotionnel pourrait être considéré comme moral s'il révèle au sujet la conformité d'une situation à une norme morale qu'il s'est fixée ou qu'il a internalisée.

Bilan

La solution minimaliste adoptée en fin de compte revient à considérer séparément chaque épisode émotionnel et lui attribuer le label de « moral » s'il apparaît dans un contexte moralement pertinent. Cela implique d'une part qu'il n'y a pas d'émotion morale à proprement parler mais uniquement des épisodes émotionnels moraux, d'autre part, que pour pouvoir les identifier, il faut au préalable que l'on se soit fait une idée de ce qu'est un contexte moralement pertinent. Contrairement à ce que pensent certains auteurs (notamment ceux qui conçoivent les émotions morales comme un critère d'identification de la moralité), il n'est donc pas possible de partir d'une analyse des émotions pour délimiter le champ de la moralité.

Pour revenir à la discussion de la section 5.2, il vaut la peine de préciser que le fait d'identifier les jugements de valeur spontanés à des réactions émotionnelles, c'est-à-dire à des épisodes émotionnels, n'impose pas de penser qu'il existe des émotions proprement morales. J'ai suggéré dans cette section qu'il n'est pas judicieux de parler d'émotions spécifiquement morales. Je ne nie pas que certaines émotions jouent un plus

[299] Ainsi il écrit : « Contrairement aux émotions non morales qui peuvent être provoquées par des événements non sociaux (par exemple, la surprise causée par le claquement d'une porte poussée par le vent ou la peur que l'on éprouve au bord d'un précipice), les émotions morales sont des réactions à des événements ou comportements posés par des sujets et ayant (ou étant perçu comme ayant) un effet sur d'autres sujets (humains et parfois non humains, y compris parfois le sujet de l'action lui-même). » (2007, pp. 113-114).

grand rôle que d'autres dans notre vie morale (par exemple la culpabilité, la honte, la colère, l'élévation ou les émotions altruistes comme la compassion et l'amour) ; mais cela n'est pas suffisant pour les considérer comme intrinsèquement morales. D'autre part, il me paraît utile de laisser la possibilité aux occurrences de n'importe quelle émotion de devenir morales en fonction du contexte dans lequel elles apparaissent. En fin de compte, pour les raisons invoquées dans cette section, je suggère que l'on ne peut pas établir une distinction morale intéressante au niveau des réactions émotionnelles (qui sont des occurrences d'émotions), c'est-à-dire au niveau des jugements de valeur spontanés : ils ne peuvent être considérés comme moraux en eux-mêmes. Il s'agit donc de trouver un autre moyen de distinguer le domaine moral du domaine non moral. C'est l'objet de la section suivante.

5.4. Une tentative de distinction entre l'activité morale et non morale

Le rejet des explications de l'évolution de la morale conçue comme objet de sélection (section 5.1) ainsi que le rejet de la notion d'émotion morale comme objet de sélection (section 5.3) concourent à donner raison à Luc FAUCHER lorsqu'il suggère qu'« après tout il est plausible de penser que le terme 'moral' est un terme désignant une 'espèce pratique', c'est-à-dire une espèce dont les membres sont rassemblés parce qu'ils possèdent quelque chose en commun qui *nous* intéresse, non parce qu'ils partagent une structure inhérente » (2007, pp. 116-117). Dans cette section, c'est précisément cette voie que je vais emprunter pour individuer la morale, c'est-à-dire pour en définir les conditions suffisantes et nécessaires. Je définirai deux critères qui me paraissent remplir ce rôle : la *recherche de fondements* et la *condition altruiste*. Le choix de ces critères découle naturellement des réflexions menées jusqu'ici sur l'altruisme psychologique (chap. 3) et sur l'activité évaluative et normative (section 5.2).

Dans ce qui suit, je commencerai par exposer brièvement ma conception de la moralité (exposé des deux critères d'individuation) avant de la développer plus en détail au moyen du tableau des normes de conduites. Suivront quelques considérations sur les implications de cette manière de concevoir de la moralité.

5.4.1. Les deux critères d'individuation de la moralité

La moralité me paraît être un cas particulier de l'activité évaluative et normative telle qu'elle a été décrite à la section 5.2. Selon le tableau affectif qui y est présenté, au cours du processus réflexif, nous nous référons à des valeurs et normes pour fonder nos jugements de valeur spontanés. Cela ne revient pas à nier la possibilité de justifier de manière minimale nos réactions émotionnelles en nous demandant simplement si elles sont pertinentes dans les circonstances particulières. Mais ce faisant, je pense que nous ne cherchons pas réellement à *fonder* nos jugements ; fonder est une forme particulière de justification qui exige le recours aux normes et aux valeurs (voir p. 207). Il me semble qu'une particularité de l'activité morale (et non uniquement de l'activité évaluative) consiste précisément dans le fait de chercher à justifier nos jugements au moyen de normes et de valeurs. Ainsi, le premier critère de la moralité est celui de la *recherche des fondements*. Cela implique que les jugements de valeur spontanés (les réactions émotionnelles) ne peuvent pas être considérés comme « moraux » ; seuls les jugements sophistiqués qui ont fait l'objet d'une réflexion fondationnelle peuvent prétendre à cette description.

La condition de la recherche des fondements est assez contraignante car elle exige une double activité de justification. Comme on vient de le voir, il faut fonder nos jugements sur des valeurs et des normes. Mais dans cette entreprise, on ne peut pas faire appel à n'importe quelles valeurs et normes. Ces dernières en particulier doivent elles-mêmes être justifiées. Nous verrons plus loin comment cela peut se faire.

Le critère du fondement n'est pas encore suffisant pour individuer la morale car les jugements peuvent être *fondés* sur des valeurs et normes non morales (par exemple les normes coutumières ou d'autorité) lesquelles peuvent être justifiées. Par exemple, je peux juger qu'il est inadmissible pour un garçon de se rendre en classe en mini jupe et fonder ce jugement sur une norme de bienséance vestimentaire, laquelle est avalisée par le directeur de l'école, une personnalité hautement vénérable. Ici, le jugement a bel et bien été l'objet d'une activité fondationnelle sans pour autant qu'il puisse être considéré comme moral.

Il nous faut donc un autre critère pour individuer la morale. A la section 4.2.2, j'ai soutenu l'idée que toute action morale est altruiste. En cela, je pense que beaucoup d'éthiciens évolutionnistes ont vu juste. Mon approche se distingue cependant de la leur en ce que ce n'est pas l'*altruisme motivationnel* (les émotions altruistes) qui me paraît pertinent pour la moralité mais plutôt l'*altruisme sophistiqué* (section 3.3.3). Ce dernier réfère aux désirs, buts et intentions des gens, à ce qu'ils considèrent comme les motifs de leurs actions ; il y a altruisme lorsque les désirs, buts et intentions ont été conçues sur la base d'une prise en considération des intérêts et du bien-être d'autrui. Ainsi, au cours du processus fondationnel de nos jugements, il faut que nous menions une réflexion de type altruiste afin que les jugements puissent être considérés comme moraux. Nous verrons que cette réflexion altruiste se fera par le biais de normes et de valeurs, si bien que seuls les jugements fondés sur des normes et valeurs qui prennent directement en considération le bien-être d'autrui peuvent être considérés comme moraux.

En résumé, la moralité naît de l'utilisation conjointe de la capacité de penser en termes évaluatifs et normatifs et de la capacité de mener des réflexions de type altruiste.[300]

Dans ce qui suit, les idées présentées dans cette section seront développées dans le détail au moyen d'une classification des différentes normes de conduite. Je commencerai par préciser ce que j'entends par *valeur* et *norme de conduite* avant de distinguer les normes morales de différents types de normes non morales. Au terme de cette analyse, nous disposerons d'une explication de la manière dont les normes peuvent être justifiées, de la place du critère altruiste dans l'analyse descriptive de la moralité et du rapport entre les normes et les valeurs. Enfin nous disposerons d'un outil conceptuel pour décider quelles situations, comportements ou assertions peuvent être considérés comme moraux.

[300] Il est important de noter ici que les deux critères de la moralité sont des conditions formelles qui ne disent rien sur le contenu des normes morales. La question de savoir si une situation ou assertion est morale ne revient pas à celle de savoir si cette situation ou assertion est justifiée ou moralement bonne. Les critères de la moralité permettent de répondre à la première question et non à la seconde (cette dernière sera traitée au chapitre 7).

5.4.2. Les valeurs et les normes de conduite

Jusqu'à maintenant, les notions de *norme* et de *valeur* ont été utilisées sans plus de détails. Mais pour comprendre le tableau des normes exposé à la section suivante, certaines précisions doivent être données.

Dans le cadre de cet ouvrage, la notion de « valeur » est utilisée à la manière des sciences sociales et ne désigne aucune réalité ontologique. Elle reflète simplement les préférences exprimées par des sujets en faveur de certaines choses (action, état de fait, objet de pensée, personne, etc.).[301] Plus précisément, je considère les valeurs conscientes (par opposition aux valeurs intuitives ; à ce propos, voir p. 204) comme des concepts abstraits (courageux, bien, mal, etc.) que l'on applique à des choses. Par exemple, lorsqu'une personne dit d'une action qu'elle est bonne, elle attribue la valeur « bien » à cette action. L'attribution d'une valeur est liée à l'expression d'une appréciation ou d'une dépréciation ; on apprécie une chose si on considère qu'elle est liée à une valeur positive, et inversement si on considère qu'elle est liée à une valeur négative.

En s'inspirant de la fameuse distinction de Bernard WILLIAMS entre concept éthique fin (*thin*) et épais (*thick*) (B. WILLIAMS 1985, p. 129), on peut diviser les valeurs en deux doubles catégories : *valeurs fines* (positives ou négatives) et *valeurs épaisses* (positives ou négatives).

Les *valeurs fines positives* paradigmatiques sont le bien, le beau, le juste ; par opposition au mal, au laid ou à l'injuste qui sont des *valeurs fines négatives*. Hors contexte, une valeur fine est vide de contenu. Ce n'est qu'une appréciation que l'on applique à une certaine catégorie d'objets.

Les valeurs *épaisses* en revanche ont un contenu. L'égalité, le plaisant, le bien-être, l'universalité (ou plutôt l'orientation universaliste), la cohérence, la neutralité affective, l'objectivité, l'altruisme, la simplicité, la vie, la coopérativité, la socialité, le courage, l'harmonie ou la santé sont des exemples de *valeurs épaisses positives* alors que l'inégalité, l'égoïsme, le disproportionné, le déplaisant, le désagréable, l'intolérance comptent parmi les *valeurs épaisses négatives*. Les valeurs épaisses possèdent à la fois un contenu descriptif assez large (c'est pourquoi on les dit *épaisses*) et une dimension appréciative en ce qu'elles font référence à une valeur fine. Dit autrement, toute valeur

[301] « In den Sozialwissenschaften werden Werte etwa als 'allgemeine, einzeln symbolisierte Gesichtspunkte des Vorziehens von Zuständen oder Ereignissen' definiert. » (LUHMANN 1987, p. 433)

épaisse entre dans la classe des objets sur lesquels porte une valeur fine. Ainsi la valeur « égalité », outre sa signification descriptive, réfère à la valeur fine du bien ou du juste, c'est-à-dire qu'elle est l'objet d'une appréciation positive. Une autre manière de concevoir le lien entre les valeurs fines et épaisses est de dire que les secondes, grâce à leur contenu descriptif, caractérisent ou spécifient les valeurs fines.

Dans la section 5.2.1.v, nous avons vu comment nous choisissons nos valeurs (essentiellement en fonction de nos sentiments) et quels facteurs influencent nos choix (biais psychologiques, influence d'autrui). Mais cela ne nous dit rien sur la question de la justification. Les valeurs sont difficilement justifiables parce que, comme nous le verrons plus loin, elles se trouvent au sommet des chaînes de justifications ; ce sont ce que j'appellerai au chapitre 7 des « éléments de base » des systèmes moraux. La manière dont on peut les justifier et les hiérarchiser relève du niveau de l'éthique normative et sera traitée dans le chapitre correspondant.

Après ces quelques réflexions sur les valeurs, venons-en aux normes. Comme beaucoup de notions clés, on en trouve les utilisations les plus variées : on parle de normes de construction, normes de mesure, normes de prudence, normes morales, normes sociales, normes d'étiquette, etc. Dans le cadre de cet ouvrage il est uniquement question de *normes de conduite*, c'est-à-dire qui portent sur le comportement des gens ; elles incluent les normes sociales, morales, coutumières, etc.

En science cognitive et en théorie des jeux (de la seconde génération), les normes de conduite sont considérées comme des entités culturelles qui peuvent être transmises d'un esprit à l'autre.[302] Dans cet ouvrage je suis cette voie en utilisant la notion de norme au sens de « contenu prescriptif d'état mental » qui peut être véhiculé par des personnes douées d'intentionnalité. Plus précisément, lorsque les gens énoncent des normes, ils produisent des formules abstraites qu'ils pensent pouvoir appliquer à une

[302] La question de la fidélité de leur transmission et de la part de reconstruction lors de la transmission est un sujet débattu (à ce propos, voir sections 1.2.2 et 1.2.3). D'autre part, beaucoup d'auteurs pensent que lorsqu'elles sont acquises par un individu, les normes induisent un certain type de comportement ; en l'occurrence, la sanction des opportunistes ou des comportements conformes à ces normes (voir sections 2.3.4 et 2.3.5). Toutefois, comme nous l'avons vu à la section 5.2.2, postuler que les normes sont motivantes en elles-mêmes est un raccourci fallacieux, même si pour d'autres raisons, les individus agissent effectivement selon les prédictions de ces auteurs.

certaine gamme de situations[303] et qui sont assorties de prescription ; elles indiquent la manière dont les gens *devraient* se comporter. A l'aspect prescriptif, il faut encore ajouter l'aspect appréciatif : les gens approuvent ou jugent légitimes les normes auxquelles ils souscrivent. Cela implique qu'ils sont disposés à les justifier ; ainsi une norme doit pouvoir faire l'objet d'une justification de la part de l'individu qui l'énonce.

Cette définition permet de ne pas confondre les normes de conduite avec les mécanismes évaluatifs intuitifs, les constantes de comportement ou les proto-normes.

Les *mécanismes évaluatifs intuitifs* ou *valeurs intuitives* (voir p. 204) réfèrent à des biais ou des mécanismes psychologiques qui induisent certaines réactions dans certaines circonstances ; ils déclenchent nos réactions émotionnelles, nos jugements de valeur spontanés. Puisque ce ne sont pas des productions conscientes, on ne peut pas les considérer comme des normes de conduite.

Les *constantes de comportement* sont des phénomènes empiriquement observables à l'échelle d'une population. On les trouve dans le monde animal. Les abeilles kamikazes par exemple piquent les intrus qui s'approchent du nid. De nature purement fréquentielle, les constantes de comportement ne sont ni prescriptives, ni des productions de l'esprit ; pour cette raison, il faut se garder de les confondre avec les normes de conduites.

On peut pressentir l'existence de *proto-normes* chez les individus dont on décèle l'acquisition par l'expérience de certains types de comportements, plus précisément les comportements dont les déviances sont régulièrement sanctionnées. Une proto-norme n'est pas le fruit d'un processus réflexif complexe mais son élaboration nécessite un certain degré de conceptualisation. Il faut que l'esprit puisse former une pensée qui établisse un lien évident entre trois éléments : i) un type de situation, ii) un comportement optimal et iii) la sanction probable des comportements déviant du comportement optimal. Par contre, l'individu capable de conceptualiser une proto-norme ne sait pas l'énoncer sous une quelconque forme langagière. Ainsi, les proto-normes se distinguent des normes de conduite en ce qu'elles ne sont ni prescriptives, ni énonçables, ni potentiellement sujettes à un processus de justification.

[303] Les philosophes approchent souvent les normes au moyen d'outils logico-sémantiques ; les normes sont conçues comme des *énoncés* d'un certain type que l'on peut transcrire en langue formelle et intégrer dans des arguments logiques. Cette approche, quoique différente, est parfaitement compatible avec la conception défendue dans cette section.

En revanche, les *lois juridiques* et les *règlements* peuvent être compris comme une classe particulière des normes de conduite.[304] Ils sont prescriptifs (ils indiquent la manière dont les gens devraient se comporter dans telle ou telle situation), peuvent être conçus dans l'esprit des gens, transmis et justifiés. Une caractéristique des lois juridiques et des règlements par rapport à d'autres formes de normes de conduite est qu'ils sont institutionnalisés, c'est-à-dire officiellement instaurés par une autorité (généralement sous forme de code écrit).

En résumé, les définitions des notions de valeur et de norme présentées dans cette section sont les suivantes. Les valeurs sont des concepts abstraits que l'on applique à des choses et qui reflètent les préférences et l'engagement des sujets en faveur de certaines choses. Les normes sont des contenus d'états mentaux qui sont énoncés par des sujets ; elles qui prescrivent le comportement qu'il faut adopter dans certaines circonstances et sont susceptibles de justification de la part des sujets qui les énoncent. Dans le cadre de ma réflexion, ces définitions me paraissent intéressantes dans la mesure où, outre le simple fait de clarifier la terminologie utilisée, elles permettront plus loin d'établir une distinction claire entre le domaine moral et les autres domaines normatifs sans faire appel de manière circulaire, à la notion même de moralité.

Le fait de comprendre les normes et les valeurs comme des productions mentales incite à adopter une approche psychologique de la normativité et de l'évaluation où l'on cherche à comprendre ce que font les gens lorsqu'ils conçoivent des normes et des valeurs conscientes, y souscrivent, les justifient, et quelles motivations à l'action leur sont corrélées. La plupart de ces questions ont déjà été traitées. Nous avons vu quelles sont les origines de la pensée normative (section 2.3.5), ce qui nous pousse à souscrire à une norme ou à une valeur (section 5.2.1.v) et le rapport entre l'adhésion à une norme ou valeur et la motivation à l'action (section 5.2.2). Ce qui nous intéressera dans les prochaines sections est la manière dont les normes peuvent être justifiées. Nous verrons que par ce biais, il sera possible de préciser le rapport entre les normes et les valeurs ainsi que le deuxième critère de la moralité : la condition altruiste. Nous verrons aussi que ce sont précisément les différentes manières dont les normes sont conçues et

[304] A ce propos, voir Georg VON WRIGHT (1963) qui traite les lois et les règles en tout genre comme des normes.

justifiées qui permettent d'établir des distinctions utiles entre différentes catégories de normes de conduites : les normes morales, d'intérêt rationnel, coutumières et d'autorité.[305]

5.4.3. Le tableau des normes de conduite

La définition de norme de conduite proposée à la section précédente s'applique à différents types de normes. Dans cette section, quatre sortes de normes seront distinguées en fonction de la manière dont les gens les justifient : les normes morales (Nm), d'intérêt rationnel (Nir), coutumières (Nc) et d'autorité (Na). Nous verrons que les normes de chacune de ces classes sont ou de premier ordre ou de second ordre. Le tableau des normes morales illustre ces différents types.

L'idée essentielle défendue dans ce qui suit est que les normes de conduite peuvent être classées dans une de ces quatre catégories en fonction de la manière dont les gens les forment et les justifient ; en eux-mêmes, les énoncés normatifs ne sont pas moraux, coutumiers, etc.[306]

[305] Notons que de par leur caractère institutionnalisé, les lois juridiques et les règlements ne peuvent entrer dans aucune de ces catégories. Dès qu'une norme morale, coutumière, etc. devient loi, c'est-à-dire dès qu'elle est officiellement instaurée, elle échappe à la dimension personnelle ; en tant que règlement écrit, elle n'est plus conçue et justifiée par des individus particuliers. Cela n'empêche pas qu'un individu puisse conceptualiser et justifier une loi sous forme de norme personnelle ; dans ce cas, la loi devient norme morale ou norme coutumière, etc. pour l'individu en question. Cette note deviendra plus claire à la lumière du tableau des normes (section suivante).

[306] Dans la littérature évolutionnaire, il existe des vues opposées à celle-ci. Certains auteurs cherchent à définir une liste de jugements et normes proprement morales. Se fondant sur les travaux de Elliot TURIEL (1993/1991 ; 1983 ; HELWIG & TURIEL 2002), Shaun NICHOLS (2004, chap. 1) par exemple propose d'établir une distinction nette entre normes et jugements *conventionnels* d'une part et *moraux* d'autre part, ces derniers étant (i) indépendants de l'autorité, (ii) généralisables et (iii) indépendants des préférences. Toutefois, je doute de l'intérêt de cette distinction (pour une critique plus circonstanciée, voir KELLY & STICH 2008 ; Nicola KNIGHT, à paraître).

Normes morales	
de 1er ordre (Nm1)	**de 2nd ordre (Nm2)**
Requièrent de la part du sujet, une **justification ultime** qui • fait référence à la **valeur fine qu'est le « bien »** (ou le « mal ») • fait directement référence aux **intérêts et bien-être d'autrui (condition altruiste)**	Requièrent de la part du sujet, une **justification** (ou une chaîne de justification) **non ultime** qui postule au moins une ou plusieurs **normes morales de 1er ordre**

Normes d'intérêt rationnel	
de 1er ordre (Nir)	**de 2nd ordre (Nir)**
Requièrent de la part du sujet, une **justification ultime** qui • fait référence à la **valeur fine qu'est le « bien personnel »** (ou le « mal personnel ») • fait directement référence à ses **propres intérêts et bien-être**	Requièrent de la part du sujet, une **justification** (ou une chaîne de justification) **non ultime** qui postule au moins une ou plusieurs **normes d'intérêt rationnel de 1er ordre**

Normes coutumières et Normes d'autorité	
de 1er ordre (Nc1 ou Na1)	**de 2nd ordre (Nc2 ou Na2)**
Requièrent de la part du sujet, une **justification dogmatique ultime** qui • fait référence à la **valeur fine qu'est le « bien »** (ou le « mal) • fait référence aux **coutumes** ou à une **autorité arbitraire**	Requièrent de la part du sujet, une **justification** (ou une chaîne de justification) **non ultime** qui postule au moins une ou plusieurs **normes coutumières ou d'autorité de 1er ordre**

Avant de commenter en détail le tableau des normes au moyen d'exemples, quelques éclaircissements terminologiques s'imposent.

Par « justification », il faut comprendre un processus qui requiert l'activité de la raison, l'affirmation de valeurs et qui est rattaché à une exigence de généralisation. La

justification est utilisée pour convaincre une personne rationnelle[307] du bien-fondé de l'objet que l'on cherche à justifier et de rallier cette personne à sa cause. On peut justifier bien des choses : une action, un jugement, le choix d'une norme, celui d'un but, une prise de position dans un certain contexte, etc. Une justification n'a pas besoin d'être très élaborée. Par exemple, en disant « Je suis contre l'avortement parce que ce n'est pas bien d'avorter », je produis une justification (même si elle est peu convaincante). A la section 5.2.1.v (p. 207), j'ai parlé de la justification des jugements de valeur, laquelle peut être ou bien minimale, ou bien de type fondationnel (lorsqu'elle fait appel à des normes et des valeurs). Dans cette section il sera question de la justification des normes elles-mêmes. Il me semble que dans ce cadre, les gens utilisent deux sortes d'arguments justificateurs : i) les *justifications ultimes* qui se fondent sur des valeurs fines, ou plus précisément qui intègrent des valeurs fines dans leurs prémisses ; ii) les *justifications non ultimes* qui ne font pas directement intervenir des valeurs fines mais intègrent dans leurs prémisses d'autres normes, lesquelles peuvent ou non être justifiées de manière ultime. Si l'on remonte la chaîne de justification des normes, on aboutit forcément à des justifications ultimes (cette distinction deviendra plus claire avec les exemples qui suivent).

D'autre part, dans le contexte de la justification, « autrui » ne réfère pas uniquement aux êtres humains qui nous entourent mais doit être compris dans un sens assez large pour englober tout individu auquel on attribue à la fois l'existence (passée, présente ou future) et des sentiments. Les entités qui peuvent entrer dans la catégorie « autrui » sont : les êtres humains (ceux qui vivent aujourd'hui tout comme ceux qui sont encore à naître), toutes sortes de divinités, esprits ou extraterrestres, les animaux.[308]

Enfin, la clause selon laquelle la justification d'une norme morale doit *faire directement référence au bien-être et intérêts d'autrui* signifie que, lors du processus de

[307] Une personne sur une île déserte n'est pas portée à justifier ses actions ; cette nécessité apparaît uniquement si on se trouve face à des êtres que l'on considère comme rationnels (ou éventuellement si on imagine être observé et jugé par un être rationnel). De même, on n'éprouve pas le besoin de se justifier devant les animaux ou les objets (à moins précisément qu'on ne leur attribue la faculté de la raison).

[308] Puisque « autrui » peut être aussi divers, lors de l'élaboration des normes, il paraît nécessaire de se demander si une hiérarchisation s'impose parmi ce divers. Et si la réponse est positive, il s'agira de justifier cette hiérarchisation.

justification d'une norme morale, la « condition altruiste »[309] doit être remplie. Cette condition exige que l'on tienne compte de l'influence des actions des sujets qui se conforment à cette norme sur la vie, le développement, l'épanouissement ou le bien-être d'autrui. De plus, de telles considérations doivent être l'élément crucial qui nous fait adhérer à la norme. Ainsi la condition altruiste n'est pas remplie si l'on tient compte des intérêts et du bien-être d'autrui de manière instrumentale (par exemple lorsqu'en fin de compte on vise son propre bien-être).

Dans les trois sections suivantes, des définitions plus formelles et un bon nombre d'exemples seront proposés afin d'illustrer les distinctions entre ces différentes formes de normes de conduite.

i. Les normes morales (Nm)

Voici quelques exemples typiques d'énoncés qui réfèrent à des Nm : « Il ne faut pas mettre la vie d'autrui en danger », « Il faut toujours agir de manière à maximiser la somme de plaisir de l'ensemble des personnes concernées par son action », « Il ne faut pas voler », « Il faut agir de manière coopérative ».

Pour qu'une norme puisse être qualifiée de Nm, il faut que le sujet la forme ou la justifie en respectant les deux conditions suivantes. i) Il doit référer aux valeurs fines que sont le bien ou le mal. ii) Il doit prendre directement en considération les intérêts et le bien-être d'autrui (c'est-à-dire remplir la condition altruiste). Ces deux conditions conjointes ne sont pas nécessaires lorsqu'il s'agit des autres types de normes.

> Exemple :
> La norme selon laquelle il ne faut pas mettre la vie d'autrui en danger, est une Nm1 si elle trouve sa justification dans le fait que le sujet considère que la vie humaine est une valeur épaisse qui caractérise la valeur fine « bien ». Dans ce cas, la condition altruiste est remplie puisque la justification prend en considération la vie humaine qui concerne tous les êtres humains sans discrimination.

En revanche, une Nm2 requiert une justification ou une chaîne de justifications non ultimes qui postulent au moins une Nm1. Ce processus consiste à montrer que la

[309] Cette condition montre que l'élaboration des normes morales exige une réflexion d'ordre altruiste, au sens sophistiqué du terme.

Nm2i à laquelle on souscrit trouve sa justification dans une Nm2ii, laquelle est justifiée par une Nm2iii et ainsi de suite jusqu'à ce que l'on parvienne à une ou plusieurs Nm1. La chaîne de justification doit, en fin de compte, aboutir à une ou plusieurs Nm1.

> Exemples :
> a) La norme selon laquelle il faut toujours tâcher de maintenir un bon équilibre psychophysique personnel, est une Nm2 si elle trouve sa justification dans le fait que les personnes qui ont trouvé cet équilibre psychophysique sont mieux disposées à mettre concrètement en pratique les Nm1 qu'elles reconnaissent.
> b) La norme qui interdit d'entretenir des phantasmes pédophiles est une Nm2 si elle est justifiée de la manière suivante : cette norme soutient la Nm2ii qui condamne les actes pédophiles (sous-entendu que l'on est convaincu que le fait d'avoir des pensées pédophiles incline le sujet à soutenir le marché de la pédophilie, voire même à passer à l'acte), laquelle soutient la Nm1 qui impose de garantir le bien-être des enfants (cette dernière se justifie par référence à la valeur épaisse « bien-être des enfants », considérée comme une caractérisation de la valeur fine « bien » et par la prise en compte des intérêts et du bien-être des enfants).

Les Nm2 ne déterminent que de manière indirecte ce qui est moral et ce qui est immoral. En effet, il me semble que la moralité d'une action est uniquement déterminée par rapport aux Nm1 qui se situent au début de la chaîne de justification. D'autre part, le caractère moral semble s'estomper en fonction de la longueur de la chaîne de justification.

> Exemples :
> a) Charles soutient qu'il est interdit de désirer le malheur d'autrui (Nm2) parce qu'il est interdit de faire souffrir autrui sans raison (Nm1). Pour lui, le fait de désirer intérieurement le malheur d'autrui est immoral par voie indirecte ; c'est immoral dans la mesure où cela incline le sujet à faire souffrir autrui sans raisons. Sur l'échelle de l'immoralité, il paraît effectivement plus pervers de faire souffrir autrui sans raison plutôt que de se contenter de nourrir de mauvaises pensées envers autrui.
> b) Imaginons Catherine, une partisane de la lutte contre le sida, qui cherche à tout prix à endiguer la transmission de la maladie. Catherine pense qu'il est moralement mauvais de prendre le risque de contaminer autrui (pour elle, la contamination est une valeur négative qui caractérise la valeur fine « mal »). Dans le but de soutenir la Nm1 à laquelle elle souscrit, elle édicte la Nm2 qui prescrit d'utiliser le préservatif. Mais il est

clair que pour Catherine, ne pas utiliser de préservatif n'est immoral que de manière héritée par le fait de prendre le risque de contaminer autrui.

Une norme devient morale uniquement par le biais de la conception que s'en font les gens. Si pour une norme, une personne peut proposer une chaîne de justification qui aboutit à une Nm1, alors on peut considérer qu'elle souscrit à une Nm2. Si pour une norme, une personne peut donner une justification ultime qui fait référence à la valeur fine « bien » ou « mal » et remplit la condition altruiste, alors on peut considérer qu'elle souscrit à une Nm1. Si pour cette même norme, une personne se refuse à donner une justification de ce type (elle peut par exemple se contenter d'en donner une justification dogmatique de type « tout le monde pense cela », ou ne pas tenir compte des intérêts et du bien-être d'autrui), alors on ne peut pas dire qu'elle souscrit à une Nm1. En bref, un même énoncé normatif, selon les circonstances dans lesquelles il est produit, peut avoir plusieurs statuts.

Exemple :

Jérôme et Xavier souscrivent à la norme suivante : « Il ne faut pas voler ».

Jérôme énonce cette norme dans le contexte suivant. Lorsqu'on lui demande pourquoi il prône cette norme, Jérôme répond qu'il ne faut pas voler parce qu'en faisant cela, on fait du tort à autrui. On poursuit le questionnement en demandant à Jérôme pourquoi il ne faut pas faire de tort à autrui. Il répond que le bien-être d'autrui est un bien qu'il faut préserver dans la mesure du possible. Ainsi, Jérôme a fondé la norme selon laquelle il ne faut pas voler sur la norme selon laquelle il ne faut pas faire du mal à autrui, laquelle se fonde sur la valeur épaisse « bien-être d'autrui » qui caractérise la valeur fine « bien ». Dans sa justification, Jérôme intègre également la condition altruiste puisque en se conformant à la norme, le sujet est censé préserver le bien-être d'autrui. Ainsi, sortant de la bouche de Jérôme, la norme suivant laquelle il ne faut pas voler est bel et bien une norme morale (plus précisément une Nm2).

Xavier, qui soutient la même norme, l'énonce dans le contexte suivant. Lorsqu'on lui demande pourquoi il prône cette norme, Xavier répond que c'est parce que tout le monde pense comme cela. Cette justification ne remplit ni la condition d'une Nm2, ni celle d'une Nm1. Xavier ne souscrit donc pas à une norme morale (en réalité, il s'agit d'une norme coutumière). Admettons qu'on demande à Xavier de réfléchir plus longuement et de donner une réponse nuancée. Il pourrait alors répondre que cette norme se défend parce que ce que pense tout le monde est ce qu'il faut penser. Dans ce cas, Xavier fonde la norme qui interdit de voler sur la norme qui impose de suivre l'avis

convergent de tout le monde. Si on poursuit le questionnement et qu'on lui demande pourquoi l'avis convergent de tout le monde est celui auquel il faut souscrire, Xavier pourrait répondre que cet avis est bon, simplement. Ainsi, Xavier pense que l'avis convergent de tout le monde est une valeur épaisse qui caractérise la valeur fine « bien ». En développant sa justification de la norme qui impose de suivre l'avis de tout le monde, Xavier s'est fondé sur la valeur fine « bien ». Mais la condition altruiste n'est pas remplie car tout le monde peut s'accorder sur des normes qui n'ont pas de rapport direct avec les intérêts et le bien-être d'autrui (par exemple, qu'il est bon de maintenir une certaine hygiène corporelle). Ainsi, on ne peut toujours pas considérer la norme souscrite par Xavier comme une Nm. Admettons que Xavier réfléchisse encore un instant avant de répondre que l'avis convergent de tout le monde est l'avis auquel il faut souscrire parce qu'il favorise la cohésion entre les membres de la société, laquelle est à l'avantage de tout le monde. Dans ce cas, la cohésion est érigée au rang des valeurs qui précisent le contenu de la valeur fine « bien » et la condition altruiste est remplie puisqu'en se conformant à la norme qui impose de suivre l'avis de tout le monde, le sujet contribue à la cohésion entre les membres de sa société. Ainsi, la chaîne de justification présentée par Xavier nous permet de considérer la norme selon laquelle il ne faut pas voler comme une norme morale (plus précisément une Nm2 qui se fonde sur une autre Nm2 laquelle se fonde sur une Nm1).

ii. Les normes d'intérêt rationnel (Nir)

Voici quelques exemples typiques d'énoncés qui réfèrent à des Nir : « Il faut éviter les excès (témérité, gourmandise, etc.) », « Il ne faut pas fumer », « Il faut manger des aliments sains », « Il faut tâcher de maîtriser ses émotions ».

Pour qu'une norme puisse être qualifiée de norme d'intérêt rationnel,[310] il faut que le sujet la forme ou la justifie en respectant les trois conditions suivantes. i) Il doit référer aux valeurs fines que sont le bien personnel ou le mal personnel. ii) Il doit prend en considération ses intérêts et bien-être personnels. iii) Il ne doit pas prendre directement en considération les intérêts et bien-être d'autrui (s'il le fait, c'est parce qu'en définitive cela lui apporte un avantage personnel).

[310] La rationalité doit être comprise ici au sens où elle est utilisée dans le contexte économique ; elle concerne la meilleure manière de maximiser les intérêts personnels.

Exemple :
La norme selon laquelle il faut éviter toute sensation douloureuse est une Nir1 si elle trouve sa justification dans le fait de considérer la douleur personnelle comme une valeur épaisse négative qui caractérise la valeur fine « mal personnel ».

En revanche, une Nir2 requiert une justification ou une chaîne de justifications non ultimes qui postulent au moins une Nir1. Ce processus consiste à montrer que la Nir2i à laquelle on souscrit trouve sa justification dans une Nir2ii, laquelle est justifiée par une Nir2iii et ainsi de suite jusqu'à ce que l'on parvienne à une ou plusieurs Nir1. La chaîne de justifications doit, en fin de compte, aboutir à une ou plusieurs Nir1.

Exemple :
La norme selon laquelle il ne faut pas prendre des risques démesurés est une Nir2 si elle est justifiée par la Nir1 qui prescrit de faire son possible pour préserver sa propre vie (laquelle est fondée sur la valeur « vie individuelle » qui caractérise la valeur fine « bien personnel »).

Il est à noter que le rayon de validité d'une Nir n'est pas forcément réduit à l'individu qui souscrit à cette norme. On peut tout à fait imaginer un individu qui édicte des Nir valables également pour autrui (voire même uniquement pour autrui !). Si c'est le cas, le sujet cherche à imposer certains comportements à autrui dans le but d'assurer son propre développement, épanouissement ou bien-être.

Exemples :
(a) La norme « il faut que tout le monde réponde au plus infime de mes désirs » est une Nir2 si elle trouve sa justification dans la Nir1 qui prescrit de chercher à réaliser le moindre de ses désirs (sous-entendu que l'assouvissement des désirs personnels est érigé au rang des valeurs épaisses qui caractérisent la valeur fine « bien personnel »).
(b) La norme qui impose à quiconque de soigner son hygiène corporelle est une Nir2 si elle trouve une justification complexe du type : i) l'hygiène corporelle permet aux gens de se sentir mieux, ii) du point de vue individuel il est très agréable d'être entouré par des gens sereins iii) « il faut rechercher l'agréable » (parce que l'agréable est une valeur épaisse qui caractérise la valeur fine « bien personnel »).

D'autre part, une Nir peut tout à fait prendre l'allure d'une Nm sans en être une.

Exemple :

Jérôme souscrit à la norme selon laquelle il ne faut pas tuer autrui. Mais s'il soutient cette norme, c'est parce qu'intérieurement, il pense qu'elle est un bon moyen pour réaliser une autre norme qui lui tient plus à cœur : celle qui impose de mettre en place les moyens nécessaires à sa sécurité personnelle (la sécurité personnelle étant érigée au rang des valeurs qui caractérisent la valeur « bien personnel »). En fin de compte, Jérôme ne fait que souscrire à une Nir.

Notons que cette analyse des normes morales repose sur un choix théorique significatif : les normes morales de premier ordre ne sont jamais directement liées au bien-être personnel. En d'autres termes, contrairement à certains philosophes (ROTTSCHAEFER & MARTINSEN 1990 ; SOBER 1993), je rejette l'idée de devoir moral envers soi-même.[311]

iii. Les normes coutumières (Nc) et les normes d'autorité (Na)

Voici quelques exemples typiques d'énoncés qui réfèrent à des Nc[312] ou Na : « Il faut porter des habits à longue manche lorsqu'on entre dans une église », « Il faut adopter une attitude polie envers autrui ».

Pour qu'une norme puisse être qualifiée de Nc ou de Na, il faut que le sujet la forme ou la justifie en respectant les deux conditions suivantes. i) Il doit référer aux valeurs fines que sont le bien ou le mal. ii) Il doit faire preuve de dogmatisme, c'est-à-dire faire uniquement référence aux us et coutumes (Nc1) ou à une autorité arbitraire (Na1) ; il ne remplit donc par la condition altruiste.

Les justifications de Nc ou Na prennent souvent les formes suivantes : « Tout le monde le fait » (Nc) ou « Cela s'est toujours fait de cette manière » (Nc) ou « X (dieu,

[311] Un argument en faveur de cette thèse a été proposé à la section 4.2.2.

[312] Il faut distinguer entre les *normes coutumières* et les *coutumes.* Les deux sont transmises par voie culturelle mais les premières sont des énoncés produits par des sujets alors que les secondes sont des régularités comportementales (envers des choses semblables dans des occasions semblables). S'il faut établir un lien entre les normes coutumières et les coutumes, il sera de l'ordre du soutien : la transmission des coutumes est renforcée par des normes coutumières.

A ce propos, notons la distinction entre une *coutume* et une *constante de comportement* (qui est une simple régularité comportementale). Au contraire de la première, la seconde ne nécessite pas la transmission culturelle ; elle comprend par exemple tous les comportements génétiquement déterminés.

la loi, le curé, le chef du village, ma mère, moi, etc.) a dit que… » (Na). D'autre part, de manière plus ou moins explicite, ces formules véhiculent l'idée que ce que tout le monde fait est bien ou ce que dit la coutume ou l'autorité est bien (c'est-à-dire que tout ce qui émane de la coutume ou de l'autorité est une valeur épaisse qui caractérise la valeur fine « bien »).

> Exemples :
> (a) Serge souscrit à la norme selon laquelle il ne faut pas courtiser la femme de son voisin. Pour justifier cette norme, il se contente de dire que cela ne se fait pas dans son pays (sous-entendu que ce qui se fait dans son pays est bien). Serge souscrit donc à une norme coutumière.
> (b) Admettons que Pierre souscrive à la même norme que Serge et lorsqu'il s'agit de la justifier, il répond qu'il ne faut pas courtiser la femme de son voisin parce que Dieu l'a ordonné et que la parole de Dieu est bonne (dans ce cas, Pierre attribue explicitement à « la parole de Dieu » une valeur épaisse qui caractérise la valeur fine « bien »). Pierre souscrit donc à une norme d'autorité.[313]

En revanche, une Nc2 ou une Na2 requiert une justification de type instrumental. Ce processus consiste à montrer que la Nc/a2i à laquelle on souscrit trouve sa justification dans une Nc/a2ii, laquelle est justifiée par une Nc/a2iii et ainsi de suite jusqu'à ce que l'on parvienne à une ou plusieurs Nc/a1. La chaîne de justification doit, en fin de compte, aboutir à une ou plusieurs Nc/a1.

> Exemple :
> Considérons la norme selon laquelle les femmes doivent porter le voile lorsqu'elles sortent de chez elles est une Nc/a2 si elle est justifiée dans le fait qu'elle a pour fonction de soutenir la Nc/a1 qui impose aux hommes de ne pas courtiser les femmes de leurs voisins.

Au vu de ces définitions, on pourrait se demander comment traiter le cas des personnes qui se contentent de fournir des justifications dogmatiques propres aux Nc ou aux Na et affirment néanmoins que les normes qu'ils justifient de cette manière sont morales. Au vu de l'analyse présentée ici, malgré ce qu'affirment ces gens, ils ne

[313] Notons que dans les deux cas, on ne peut pas parler de Nm1 puisque la prise en compte des intérêts et du bien-être d'autrui (condition altruiste) n'intervient pas dans la justification.

souscrivent à rien de plus qu'à des Nc/a, puisqu'une norme ne devient morale qu'à partir du moment où l'on est disposé à en donner une justification ultime qui tienne compte des intérêts et du bien-être d'autrui.

Notons qu'une Nc/a peut devenir une Nm dès lors que le sujet est prêt à en donner une justification propre aux Nm.

> Exemple :
> En 1999, Myriam affirmait qu'il ne faut pas courtiser la femme de son voisin « parce que Dieu l'a dit et que la parole de Dieu est bonne ». Voilà une justification dogmatique et de ce fait, en t_1, Jérôme se contentait de justifier une Na.
> Aujourd'hui, Myriam, affirme qu'il ne faut pas courtiser la femme de son voisin parce que Dieu l'a dit *et* qu'il est le seul garant du bien de l'humanité. Ici, Myriam sous-entend que le bien de l'humanité est bon (en d'autres termes, elle affirme que le bien de l'humanité est une valeur épaisse qui caractérise la valeur fine « bien »). D'autre part, la notion de bien de l'humanité implique autrui si bien que dans le processus de justification Myriam prend en en considération l'influence des actions du sujet qui suit les impératifs divins sur la vie, le développement, l'épanouissement ou le bien-être d'autrui. Ainsi, on peut dire qu'en t_2, Myriam souscrit à une Nm.

De même, une Nc/a peut devenir une Nir dès lors que le sujet est prêt à en donner une justification propre aux Nir.

> Exemple :
> En 1999, Rémi souscrivait à la norme selon laquelle il ne faut pas courtiser la femme de son voisin parce que son père le lui a souvent répété. Ainsi en t_1, Rémi justifiait une Na.
> Aujourd'hui, Rémi affirme qu'il ne faut pas courtiser la femme de son voisin parce qu'il est marié et qu'il a beaucoup de voisins. Dans ce cas, en t_2, Rémi souscrit à une Nir2 qui a se fonde sur la Nir1 selon laquelle il faut mettre en place tout les garde-fous nécessaires au maintien de ses avoirs personnels (sous-entendu que les avoirs personnels sont des valeurs qui caractérisent la valeur fine « bien personnel »).

En résumé, le type des normes (morales, d'intérêt rationnel, coutumières ou d'autorité) est à la fois relatif à l'agent et au temps.

iv. Quelques précisions

Au terme de cette catégorisation des différents types de normes j'apporte deux précisions. Premièrement, même si les différentes définitions proposées ci-dessus permettent de catégoriser aisément à peu près tous les énoncés normatifs, il restera toujours des cas limite.

> Exemple :
> Georges, un fervent chrétien, pense qu'il faut s'abstenir de tout rapport sexuel qui ne soit pas accompli à la fois dans le cadre du mariage et en vue de la procréation. Georges souscrit à cette norme parce qu'il ne veut pas blesser son Dieu qui lui a généreusement insufflé un quota d'énergie vitale avec la mention : « ne pas gaspiller ».
> Georges, souscrit-il une Nm ? A priori, il faudrait répondre que non puisque la justification de la norme selon laquelle il faut s'abstenir de tout rapport sexuel accompli hors mariage et sans but de procréation ne remplit pas la condition altruiste. Mais d'un autre point de vue, il peut s'agir d'un cas limite car il existe la possibilité de considérer que non seulement les êtres humains, mais également Dieu entrent dans la catégorie « autrui ». Or dans sa justification, Georges considère le bien-être de Dieu (il ne veut pas le blesser).

Deuxièmement, toute norme qui prescrit un comportement en rapport avec la survie, le développement l'épanouissement ou le bien-être d'autrui et pour laquelle on donne une justification ultime n'est par forcément une Nm. Ce qui est décisif pour que l'on puisse parler de Nm, c'est que la justification de la norme prenne *directement* en compte les intérêts et le bien-être d'autrui.

> Exemples :
> (a) Imaginons un petit village de montagne dans lequel tous les hommes portent des pantalons. Imaginons qu'un jour, un membre de ce village (Julien) décide de porter une jupe. Il est vrai qu'il existe des régions où les hommes portent régulièrement des jupes, mais dans ce village, personne n'a jamais vu un homme dans une telle tenue vestimentaire. Un beau jour, Julien décide de porter cet habit pour se rendre à la messe dominicale. A sa vue, les anciens du village sont outrés, tonnent que cela ne se fait pas et qu'il faut respecter la bonne vieille tradition ; ils se sentent personnellement agressés. La décision de Julien de porter une jupe en ce dimanche matin a heurté une norme à laquelle souscrivent les anciens (« Les hommes doivent porter des pantalons ») et leur a

causé un tort psychologique. Doit-on en conclure que la norme souscrite par les anciens est une Nm ? Il est vrai que l'action de Julien a égratigné le bien-être des anciens du village en leur causant un tort psychologique. Il est vrai que l'action de Julien contredit une norme à laquelle souscrivent les anciens du village. Malgré cela, je pense que le seul tort qui peut être imputé à Julien est celui de contredire une Nc2 (« Les hommes ne doivent pas porter de jupe »), laquelle repose sur une Nc1 (« Il faut respecter la tradition »), laquelle ne remplit pas la condition altruiste.

(b) De même, la norme qui prescrit de ne pas tuer autrui n'est qu'une Na si elle est justifiée par le seul recours à l'autorité divine. Dans ce cas, il est vrai que la norme prescrit un comportement spécifique à adopter vis-à-vis d'autrui. Par contre, les intérêts et le bien-être d'autrui ne sont pas pris en considération dans la justification de cette norme.

On pourrait objecter que la conception des Nm présentée ici est trop restreinte pour correspondre au sens commun. Il est vrai que beaucoup d'énoncés qui semblent être des normes morales peuvent, selon les circonstances, ne pas en être ; tout dépend de la manière dont on les intègre et les justifie. Mais cela ne me paraît pas choquant. Considérons un enfant auquel les parents enseignent de ne pas voler. Selon le modèle proposé ci-dessus, il intègrera probablement cette norme sur le mode de l'autorité ou de l'intérêt rationnel jusqu'à ce qu'il soit capable de comprendre que l'application de cette norme permet de ne pas léser autrui. Pourrait-on dire qu'un enfant qui applique cette norme en pensant éviter la punition agit de manière morale ? Il semble précisément que le sens commun s'accorde avec l'idée que cet enfant agit simplement *conformément à* la norme que ses parents tentent de lui inculquer sans pour autant que son action puisse être considérée comme morale. Ainsi, je ne pense pas que l'approche présentée ici manque son rendez-vous avec le sens commun.

5.4.4. D'autres concepts moraux

Nous avons vu que les normes morales se distinguent des normes non morales en ce qu'elles remplissent le critère de la condition altruiste. Mais cette condition agit uniquement au niveau du choix et de la justification des normes. Partant de là, c'est en

quelque sorte le rayonnement des normes morales qui confère à d'autres choses le caractère de la moralité.[314]

Ainsi un jugement peut être considéré comme moral s'il se fonde sur une norme morale. Un jugement ne peut donc pas être moral en lui-même, indépendamment du contexte. Par exemple, dire « Charles est courageux ! » n'est un énoncé moral que si l'on souscrit à la norme morale selon laquelle il faut être courageux. Quant à l'intention, elle est morale si elle projette la réalisation d'une action conforme à une norme morale. Enfin une action morale est une action produite en conformité avec une intention morale.

Notons que selon cette dernière définition, si une action est produite conformément à une norme morale mais non à l'intention morale d'agir conformément à cette norme, alors on ne peut pas parler d'action *morale*.

> Exemple :
> Léon défend la Nm selon laquelle il faut venir en aide aux personnes en détresse et la Nir selon laquelle il faut réaliser les actions qui sont socialement reconnues. Un matin, lors de sa promenade quotidienne le long des quais bondés du port du village, il voit un enfant se noyer ; mais avant de se lancer à l'eau pour sauver l'enfant, il forme l'intention de suivre la Nir et non la Nm. Dans cette situation, Léon a agi conformément à la Nm sans que son action puisse être considérée comme morale.

D'autre part, toutes les actions morales n'ont pas forcément comme effet de satisfaire une norme morale. Etant donné que le critère de la moralité d'une action morale se trouve dans l'intention morale, il se peut que l'intention échoue à réaliser la prescription d'une norme morale.

> Exemple :
> Dominique souscrit à la Nm selon laquelle il faut soigner les animaux et ne pas leur faire de tort. Un jour, en se promenant, elle voit un oisillon par terre, au pied d'un arbre, qui piaille à tue-tête. En observant la situation, Dominique constate que cet oisillon est tombé d'un nid qui se trouve dans l'arbre, au deuxième niveau, troisième branche à gauche. Dans un souci de bienfaisance et pour appliquer sa Nm, Dominique forme l'intention de sauver l'animal. Elle prend délicatement l'oisillon dans ses mains, grimpe

[314] Les distinctions qui suivent concernent les normes morales, mais des réflexions similaires peuvent être faites au sujet des autres types de normes de conduite.

> dans l'arbre et le replace dans le nid. Ce que Dominique ne sait pas, c'est que la mère de l'oisillon s'était habituée à le nourrir par terre depuis qu'il était tombé du nid et qu'en le retrouvant soudain dans le nid, imprégné d'une répugnante odeur humaine, elle ne le reconnaîtra plus et le laissera mourir de faim. On peut dire de l'action de Dominique, qu'il s'agit d'une action morale qui a échoué à réaliser l'intention morale.

La définition proposée pour les actions morales est très restrictive car elle dépend de la formation d'une intention consciente préalable à l'action. Or il arrive souvent que l'on agisse sans réfléchir auparavant.[315]

> Exemple :
> Vera, une riche dame, rencontre un pauvre vieillard souffrant dans la rue. Cette vision la trouble à tel point que sous le coup d'un fort sentiment de compassion et sans réfléchir une seconde, elle lui offre tout l'argent et les bijoux qu'elle porte sur elle.
> L'action de Vera ne peut pas être qualifiée de morale puisqu'elle n'est pas le résultat d'une intention de se conformer à une Nm ; elle est simplement le résultat d'un puissant sentiment de sympathie. Par contre cette action est conforme à la norme qui exige d'aider son prochain.

Ainsi un grand nombre d'actions peuvent n'être que *conformes aux normes morales* ; elles réalisent les prescriptions imposées par des normes morales sans pour autant être corrélées dans la pensée des agents à l'intention de suivre ces normes. Certains lecteurs y verront une difficulté et j'avoue que sur ce point, ma théorie s'écarte quelque peu du sens commun. Je pense cependant qu'il ne faut pas dénigrer l'importance des actions *conformes aux normes morales acceptées*, car la plupart d'entre elles résultent d'habitudes de comportements parfois acquises au prix d'importants efforts d'apprentissage social, de recherche de cohérence, de discussions et interactions morales avec son entourage (avec toutes les influences mutuelles que cela comporte).[316] D'autre part, une action conforme aux normes morales, même si elle n'est pas morale de par la manière dont elle a été produite, peut parfaitement être *moralement pertinente* pour un observateur qui la condamne ou la loue (selon les normes et les valeurs qu'il défend lui-même).

[315] Sans mentionner le fait que l'intention consciente n'est sans doute pas causalement responsable de la motivation à l'action (à ce propos, voir la section 5.2.2 sur la motivation à l'action).

[316] Pour un point de vue similaire, voir GIBBARD 2002/1990.

Le cas des valeurs conscientes est un peu plus complexe. De même que les normes, jugements, les intentions et les actions, elles ne sont pas morales en elles-mêmes, c'est-à-dire indépendamment du contexte dans lequel elles sugissent. En revanche, selon les circonstances, elles peuvent le devenir de manière directe ou indirecte. Les valeurs peuvent être considérées comme morales si elles sont utilisées dans le cadre de justifications de normes morales ; ainsi, de même que les intentions, les jugements et les actions, elles acquièrent la moralité par effet de répercussion des normes morales.[317] Les valeurs peuvent également être considérées comme morales ou non selon la manière dont elles sont conçues par les sujets ; si leur conception résulte de la prise en considération des intérêts et du bien-être d'autrui, alors elles sont morales. Ainsi, ce ne sont pas uniquement les normes mais également les valeurs conscientes qui peuvent remplir la condition altruiste si bien que peuvent être moraux, non seulement les jugements normatifs, mais également les jugements de valeur (ceux qui se fondent uniquement sur des valeurs).

5.5. La morale comme produit dérivé

Au début de ce chapitre (section 5.1), j'ai avancé l'idée que la moralité est un effet dérivé non adaptatif qui « surfe » sur d'autres traits adaptatifs qui ont évolué de manière propre. Il en va de même pour l'activité évaluative et normative en général. Cela tient à ce que ces deux phénomènes sont très similaires et ne se différencient qu'en fonction de critères externes. Le premier critère d'individuation de la moralité est celui de la recherche des fondements ; il impose à l'agent moral de chercher à justifier ses jugements au moyen de normes et de valeurs. Le second critère, la condition altruiste, exige que dans le cadre de cette justification, l'agent moral prenne en considération les intérêts et le bien-être d'autrui. Dit autrement, les normes et valeurs morales doivent être conçues sur la base d'une réflexion sur ce qui est bien pour autrui. Ces deux critères ont l'avantage de définir la moralité de manière non circulaire.

[317] Ainsi, la même valeur, selon qu'elle est invoquée dans le contexte de la justification d'une norme ou d'une coutume, pourra tantôt être considérée comme morale tantôt non.

A la lumière de l'ensemble des réflexions menées sur l'activité évaluative en général, le rôle des émotions en morale ainsi que sur les critères d'individuation de la moralité, les bases sont posées pour déterminer sur quelles capacités cognitives et biais psychologiques l'activité morale repose.[318]

Pour commencer, la condition altruiste exige que les individus soient capables de faire la distinction entre eux-mêmes et autrui et de comprendre les besoins et désirs d'autrui. A cet effet, il faut disposer d'une forme de conscience de soi et de la capacité de la théorie de l'esprit (voir p. 137).

La capacité de penser en termes évaluatifs et normatifs dont nous avons vu à la section 2.3.5 qu'elle est évolutionnairement avantageuse, repose sur un certain nombre de traits adaptatifs. Parmi ceux-ci, il y a les émotions ; en effet, le tableau affectif illustre le fait que les réactions émotionnelles sont à la base du processus réflexif évaluatif. Il y a également la capacité de formuler et appliquer les concepts et énoncés abstraits que sont les valeurs et les normes ; cette capacité repose sur une compréhension des avantages (ou désavantages) liés à certains comportements ou à la présence de certains traits. De telles activités nécessitent la possession d'une mémoire autobiographique (qui permet le rappel des événements de la vie antérieure) ainsi que la possibilité de former des méta-représentations, c'est-à-dire des représentations qui portent sur d'autres représentations.

Pour pouvoir entrer dans des discussions normatives et influencer mutuellement nos choix de normes et de valeurs ainsi que nos réactions émotionnelles, il faut faire entrer en jeu la transmission culturelle qui repose sur l'apprentissage social et en particulier l'imitation, comprise au sens lâche de reconstruction plus ou moins fidèle des entités culturelles observées (section 1.2.1). D'autre part, il vaut la peine de mettre l'accent sur le fait que l'utilisation du langage renforce l'efficacité de l'apprentissage social. Robin DUNBAR (1996) a développé une thèse assez crédible selon laquelle le langage a évolué en premier ressort pour remplir le besoin de commérage ; il permettait aux gens de savoir ou de se souvenir de qui a fait quoi, qui est où, qui est de confiance, etc. Puis, la combinaison du langage et d'une forme évoluée de théorie de l'esprit a

[318] La définition formelle de la morale proposée dans cet ouvrage est très exigeante. Si l'on préfère en revanche travailler avec une définition plus lâche (c'est envisageable puisque la morale est une espèce pratique » ; voir p. 237), il faudra probablement assouplir également la liste des capacités et biais psychologiques sur lesquelles repose la moralité.

permis à des grands groupes d'individus non parents de bénéficier des effets de la coopération.

Quant au pouvoir de justifier les normes, il repose sur une capacité de réflexion (ou intelligence) hautement développée. Cette dernière a certainement été sélectionnée en partie pour mieux gérer les environnements changeants auxquels les êtres humains ont constamment dû se réadapter depuis le pléistocène (section 1.2.1) ainsi que sous l'effet de la course aux armements définie par TRIVERS (1971 ; voir section 3.4.1), sa fonction principale étant de maximiser la réalisation des buts, plus précisément d'évaluer si oui ou non une action permet d'atteindre des buts donnés. De même que pour la transmission culturelle, la capacité de réflexion a pu se complexifier et s'affiner de manière impressionnante grâce à l'évolution du langage. Une fois développée, l'intelligence a pu être utilisée à d'autres effets, en l'occurrence dans un contexte d'activité normative (car elle mène les hommes à prendre conscience de la nécessité de suivre certaines règles; voir section 2.3.5), puis de justification morale (SINGER 1981).

Le tableau affectif de la motivation (section 5.2.2) a montré que la motivation à agir conformément aux normes et valeurs auxquelles on adhère, repose entièrement sur les sentiments et en particulier les réactions émotionnelles. Dans la section 5.3.3, une liste des émotions les plus couramment déclenchées face aux situations moralement pertinentes a pu être établie ; il s'agit des émotions empathiques (voir section 3.4.6) comme la sympathie ou la compassion[319] qui fondent généralement la motivation altruiste et nous incitent à agir conformément aux normes morales (puisque ces dernières sont précisément de nature altruiste) ; il s'agit également des émotions liées à la punition (colère, dégoût moral)[320] qui incitent à réprimander les comportements déviant des normes. Il doit être possible d'imaginer des scénarios évolutionnaires pour les émotions simples comme la sympathie ou la colère ;[321] les émotions complexes comme la culpabilité ou la honte nécessitent en revanche une analyse plus circonstanciée qui tient compte des contingences culturelles (voir DEONNA 2007).

Les traits listés ci-dessus sont suffisants pour que l'activité morale puisse avoir lieu. Il existe toutefois d'autres traits sur lesquels repose la moralité et qu'il faut se

[319] Parmi les complémentaires des émotions empathiques, on peut compter la gratitude ou l'élévation.

[320] Parmi les complémentaires des émotions punitives, on peut compter la culpabilité ou la honte.

[321] Pour des explications de l'émergence, de l'origine et du fonctionnement d'un certain nombre d'émotions sociales, voir HAIDT 2003 et FESSLER & HALEY 2003.

garder de négliger. Il s'agit de deux types de biais psychologiques (biais de transmission et de contenu ; voir section 1.2.3) qui influencent le contenu même des valeurs et normes morales.

Les biais de transmission les plus importants sont la tendance à ce conformer à la majorité et celle à imiter certains modèles comme les membres de la famille ou les individus prestigieux (p. 43 ; section 5.2.1.v).

Parmi les biais de contenu, on compte les émotions. Grâce au fait que l'activité morale repose généralement sur un groupe restreint d'émotions (section 5.3.3) qui présentent des réactions typiques face à certains types de situations, une objectivité de fait peut être garantie ; en effet, nous avons vu aux sections 5.2.1.v et 5.2.1.vi que le choix des normes morales est largement influencé par les réactions émotionnelles. D'autres biais de contenu qui influencent nos réactions émotionnelles sont les valeurs intuitives dont il était notamment question à la section 5.2.1.iii : l'aversion face à la souffrance d'autrui (NICHOLS 2004) et en particulier la souffrance causée de manière intentionnelle (p. 204), l'aversion face à l'iniquité (FEHR & ROCKENBACH 2003)[322], le dégoût face à l'inceste (p. 233), etc.[323]

Au terme de cette analyse, force est de constater que la moralité est cognitivement très exigeante ; elle peut uniquement être pratiquée par des individus dotés de capacités cognitives plus ou moins complexes. Dès lors qu'un individu possède toutes les capacités essentielles à l'activité morale, on pourra dire de lui qu'il est capable de moralité. Les individus qui ne possèdent pas l'ensemble de ces capacités ne méritent pas réellement le qualificatif d'*agents moraux* et doivent être considérés comme des êtres fondamentalement amoraux. [324]

Un certains nombre de traits adaptatifs mentionnés dans cette section se trouvent dans le monde animal et chez les petits enfants. Certaines espèces d'animaux sociaux non humains ainsi que les petits enfants possèdent la conscience de soi, la théorie de

[322] Notons que selon certains auteurs, on n'acquiert pas automatiquement la valeur intuitive de l'équité ; elle ne serait donc pas génétiquement déterminée ou encapsulée dans un module. On possèderait par contre une propension à l'acquérir (OSTROM 1998).

[323] Pour d'autres exemples, voir note 244; voir aussi HARMAN (1999) et HAIDT & CRAIG (2004).

[324] En revanche, les normes morales peuvent tout à fait s'appliquer aux êtres amoraux puisque, par définition, elles tiennent compte du bien-être d'autrui, c'est-à-dire de tout être doué de sentiments.

l'esprit et l'empathie.[325] On peut trouver chez les animaux certains biais de contenu qui influencent nos jugements moraux et le choix de nos valeurs et normes ; l'aversion face à la souffrance d'autrui[326] et l'aversion de l'iniquité[327] sont attestées chez certains singes. Par contre, on peut sérieusement douter de la présence de certaines capacités chez les petits d'homme et les animaux. La mémoire autobiographique et la conceptualisation de méta-représentations par exemple semblent n'émerger chez les êtres humains que vers l'âge de quatre ans (BARTH *et al.* 2004 ; voir aussi CLEMENT 2007). Quant aux animaux, on sait qu'ils sont capables de représentations mentales concrètes (un chien par exemple peut se faire une représentation mentale de l'os et saura si ce que lui montre son maître est un os ou une poupée), mais il semblerait qu'ils soient incapables de former des représentations mentales abstraites comme des normes ou des valeurs (PROUST 2003).[328] De plus, beaucoup de penseurs leur refusent la faculté d'imiter, nécessaire à la transmission culturelle (voir section 1.2.1). Enfin, il est clair que les animaux et petits enfants ne sont pas à même de se lancer dans un processus de justification. Pour ces raisons, il faut les considérer comme des êtres amoraux. En revanche, ils peuvent être candidats à la proto-moralité.[329] Cette dernière est moins exigeante que la moralité : Il suffit qu'un individu soit capable d'apprendre et gérer les interdits et comportements punitifs d'autrui et d'appliquer lui-même des comportements punitifs liés à des interdits. Les chimpanzés par exemple sont clairement capables

[325] Il semblerait que les grands singes soient capables de réactions émotionnelles assez sophistiquées qui nécessitent l'empathie cognitive (DE WAAL 2002, pp. 18-19).

[326] Jules MASSERMAN et ses collègues (1964) ont montré que des singes rhésus cessent de se nourrir s'ils comprennent que le procédé par lequel ils obtiennent leur nourriture cause une souffrance à un de leurs congénères. Frans DE WAAL (1989, 1997/1996) a pu observer que les comportements de réconciliation chez les primates sont courants et fonctionnellement importants au niveau social. On trouve même des exemples où des individus neutres viennent apportent leur assistance dans les processus de réconciliation ; ils calment les protagonistes et consolent les victimes.

[327] Sarah BROSNAN et Frans DE WAAL (2003) on montré que les singes capucins expriment une aversion face aux partages inégaux (mais uniquement lorsque l'iniquité les concerne) ; ils cessent de coopérer avec l'expérimentateur lorsqu'ils remarquent que leurs voisins reçoivent une meilleure nourriture qu'eux en échange du même service (voir aussi FLOWER *et al.* 2004).

[328] Dans le monde animal on trouve de nombreux exemples d'interdits et des comportements punitifs. Mais en ne peut pas en inférer que les animaux agissent en fonction de normes morales ; la punition peut être uniquement une réaction contre celui qui montre certains comportements sans qu'il y ait recours à une quelconque représentation normative.

[329] Notons que pour être rigoureux, il faudrait plutôt parler de proto-normativité que de proto-moralité.

d'apprendre et d'appliquer par expérience les codes de conduite utilisés par les individus de leur entourage (DE WAAL 1997/1996). La proto-moralité reflète le fait qu'il existe une continuité entre les hommes et les animaux (DE WAAL 1997/1996, BOEHM 1999 ; 2002/2000). Ce qui les distingue est une affaire de degré et non de nature, si bien qu'il n'est pas exclu qu'une espèce animale sociale autre que l'homme puisse, au fil de l'évolution, devenir capable de moralité.

5.6. Quelques implications aux niveaux métaéthique et normatif

Au terme de ce chapitre, il vaut la peine de s'arrêter un instant sur les implications, aux niveaux de la métaéthique et de l'éthique normative, des thèses qui viennent d'être défendues en éthique descriptive.

L'analyse descriptive de la moralité présentée ici implique au niveau métaéthique un certain nombre de parti pris ontologiques ou épistémiques. Nous avons vu que rien ne possède la caractéristique de la moralité indépendamment du contexte. Pour que quelque chose puisse être considéré comme moral, il faut la justification d'un sujet rationnel ainsi que la réalisation de la condition altruiste. Puisqu'elle est en quelque sorte créée par les sujets, la moralité ne possède pas une existence indépendante de ses sujets. Ainsi les positions métaéthiques réalistes morales qui contredisent précisément cela sont à rejeter.

Ce rejet implique-t-il que nous ne puissions pas prétendre à la connaissance morale ? En un sens il faut l'admettre ; nous verrons au chapitre 7 que l'on ne peut pas parler de connaissance sur le modèle empirique. Cela dit, l'objectivité de fait et l'objectivité psychologique que l'on trouve dans l'activité morale (section 5.2.1.vi) fournissent des moyens d'entente entre les différents agents moraux. En ce sens, on peut parler d'une forme minimale de connaissance morale. Une autre position métaéthique qui entre en contradiction avec mon analyse de la moralité est l'émotivisme. Selon cette position, en énonçant un jugement moral, les gens ne font qu'exprimer un sentiment d'approbation ou de désapprobation associé à une prescription (AYER 1946/1936 ; STEVENSON 1937). Or nous avons vu dans ce chapitre que ce n'est pas simplement en exprimant une émotion que l'on souscrit à des valeurs conscientes ou des normes (condition nécessaire de la moralité). Selon le tableau affectif, ce que les émotivistes appellent « jugements moraux » ne sont rien de plus qu'une manière d'exprimer nos

réactions émotionnelles (voir p. 206). Ainsi l'émotivisme n'est pas compatible avec l'analyse descriptive de la moralité présentée dans ce chapitre. Ces questions de métaéthique seront reprises et développées dans le chapitre suivant.

Le contenu de ce chapitre est d'ordre strictement *descriptif* ; l'évaluation, la normativité et la prescriptivité n'ont pas été abordées sous l'angle de ce qu'il *faut* faire ou ne pas faire. Mais cela n'empêche pas que les critères de la moralité présentés plus haut permettent de rejeter d'emblée certains courants présents en philosophie morale. Par exemple, les versions fortes de l'hédonisme qui définissent le bien en termes de bonheur personnel sont écartées de la liste des théories morales dans la mesure où elles prescrivent uniquement des normes d'intérêt rationnel (Nir), lesquelles ne remplissent pas la condition altruiste.

D'autre part, ma définition de l'action morale (section 5.4.4) s'accorde assez mal avec les théories conséquentialistes radicales qui ne prennent pas en considération l'intention des sujets lorsqu'il s'agit de juger si une action est moralement bonne ou mauvaise.[330] En effet, pour peu que l'on admette ma définition de l'action morale, on est tenté de dire que si une personne forme une intention moralement bonne et accomplit ensuite une action conforme à cette intention, alors elle produit une action moralement bonne, même dans les cas où l'action échoue. Or, c'est une idée qu'il faut abandonner si l'on défend une position conséquentialiste qui focalise l'attention sur les résultats de l'action. L'utilitarisme des actes par exemple défend le principe selon lequel il faut toujours agir de manière à maximiser le plaisir (BENTHAM 1948/1789, IV) ou le bonheur (MILL 1998/1861) de l'ensemble des individus concernés par l'action. Dans ce cadre de pensée, les actions qui échouent à réaliser l'intention morale peuvent difficilement être considérées comme moralement bonnes. De manière générale il me semble que l'utilitarisme des actes est contradictoire dans les exigences qu'il impose aux sujets. D'une part, il demande que les agents moraux forment des intentions d'agir conformes au principe utilitariste et agissent en fonction de ces intentions. Mais d'autre part, lorsqu'il s'agit de juger la qualité morale de ces actions, les intentions perdent soudain de leur intérêt au profit des conséquences. A mon avis, il s'agit ici d'une

[330] Précisons que ma définition de la moralité et ses implications n'est pas paradigmatique de l'éthique évolutionniste ; un certain nombre d'éthiciens évolutionnistes négligent la réflexion et les intentions en faveur des sentiments et des conséquences des actions. Ce genre d'approche est parfaitement compatible avec l'utilitarisme (voir par exemple RUSE 1998/1986, p. 209, p. 235 ; AGAR 2002).

inconséquence pratique qui peut mener à des situations extravagantes. Considérons un exemple : Admettons que Mélanie, une utilitariste convaincue, forme une intention d'agir conforme au principe utilitariste, mais que l'action qui découle de cette intention échoue à réaliser la plus grande somme possible de plaisir (ou bonheur). En revanche Hermine, qui se trouve à peu près dans la même situation que Mélanie, se moque de la norme utilitariste et forme ouvertement une intention d'agir qui vise uniquement à maximiser ses intérêts personnels. Or, il se trouve que l'action qui découle de cette intention égoïste a comme effet secondaire inattendu de réaliser la plus grande somme de plaisir (ou bonheur) possible. Si elle veut rester cohérente avec ses convictions utilitaristes, Mélanie devrait juger que l'action d'Hermine est moralement meilleure que la sienne en dépit des intentions égoïstes qui en sont la source. Du point de vue théorique, c'est une conclusion que l'on peut tirer mais il est certain que Mélanie sera incapable de se convaincre *sincèrement* que l'action d'Hermine est moralement meilleure que la sienne. Ainsi la position utilitariste n'est pas soutenable du point de vue de la psychologie des agents moraux.[331]

On pourrait se demander si le fait que mon analyse de la moralité implique le rejet d'un certain nombre de positions classiques en philosophie morale (réalisme moral, émotivisme, certaines formes de conséquentialisme, hédonisme individualiste) ne constitue pas un défaut gênant. Je ne le pense pas car les théories rejetées ont déjà été l'objet de puissantes critiques provenant de tous bords si bien qu'elles trouvent peu de défenseurs contemporains. Au-delà du simple rejet de certains courants philosophiques, je tâcherai de montrer au chapitre 7 que les données empiriques présentées dans ce chapitre fournissent des pistes pour fonder ou justifier des valeurs et principes fondamentaux.

Conclusion

Ce chapitre avait pour objectifs principaux de montrer que la moralité est un produit dérivé, de définir les traits adaptatifs sur lesquels elle repose et de préciser les rôles respectifs des processus cognitifs et affectifs dans son activité.

[331] Une manière de sauver l'utilitarisme de cette objection serait d'en affaiblir l'aspect conséquentialiste en accordant une certaine importance aux intentions des agents ; c'est vers cette voie que se dirige l'utilitarisme des règles (R. BRANDT 1959).

Le tableau affectif présente une double image de l'activité évaluative et normative ; dans son aspect *spontané*, elle est composée de réactions émotionnelles guidées par des valeurs intuitives ; dans son aspect *réflexif*, elle se décline en jugements sophistiqués, normes et valeurs conscientes.

C'est au niveau spontané qu'intervient la motivation à l'action. Elle provient d'une seule et unique source : les sentiments sous leur forme pure ou lorsqu'ils sont intégrés dans une réaction émotionnelle. Si nous sommes enclins à penser que les normes, valeurs et jugements sophistiqués auxquels nous adhérons sont motivants, cela tient au fait que la plupart d'entre eux sont intiment corrélées à nos réactions émotionnelles.

C'est au niveau réflexif que l'on peut trouver le grain spécifique de la moralité. Cette dernière se caractérise par une recherche de fondement de nos jugements, mais c'est essentiellement à travers le processus de justification des normes et valeurs conscientes que l'on peut l'individuer. Pour pouvoir être qualifiée de morale, une norme ou une valeur exige une réflexion d'ordre altruiste (au sens sophistiqué du terme) ; partant de là, c'est en quelque sorte le rayonnement des normes (et valeurs) morales qui confère à d'autres choses le caractère de la moralité.

Plus précisément, les deux critères de la moralité s'appliquent différemment : le critère de la recherche des fondements agit au niveau des jugements et des normes (ils doivent être fondés) alors que la condition altruiste concerne les normes et les valeurs conscientes (elles doivent être de nature altruiste). De plus, à y regarder de plus près, on constate que la recherche des fondements est subordonnée à la condition altruiste ; en effet, lorsque cette dernière est réalisée, les jugements qui en découlent remplissent forcément le premier critère de la moralité.

En fin de compte, il ressort que l'*activité morale* est largement influencée, via les réactions émotionnelles, par certaines émotions sociales (dont nous avons établi une liste plus ou moins exhaustive). En revanche, la *moralité* en tant que phénomène spécifique est relative à la manière dont des individus dotés de capacités cognitives hautement développées (il semblerait que seuls les êtres humains répondent à ce critère) conçoivent et justifient les jugements, valeurs et normes auxquels ils adhèrent. En un sens, on peut dire que la moralité n'a aucune réalité externe à cette activité réflexive. Nous verrons au chapitre suivant que cette analyse n'est pas sans conséquence sur le traitement des questions de métaéthique.

6. Métaéthique et pensée évolutionnaire

Le chapitre précédent portait sur l'éthique descriptive. J'y ai développé une analyse psychologique de la moralité en proposant un tableau affectif qui rend compte de la manière dont les êtres humains réagissent aux situations moralement pertinentes et en viennent à élaborer certaines valeurs et normes plutôt que d'autres. J'ai également défini de manière assez procédurale quelles productions évaluatives et normatives des sujets humains peuvent être considérées comme morales. Mais cette analyse descriptive ne nous fournit aucun moyen de distinguer ce qui est moralement bon de ce qui est moralement mauvais. A cet effet, il est indispensable de bénéficier d'une théorie morale normative et non seulement descriptive. La manière la plus simple et la plus tentante de justifier nos normes est de reproduire en morale le modèle de la connaissance empirique. Cette stratégie repose sur une théorie métaéthique réaliste cognitiviste. C'est de cette dernière dont il sera essentiellement question dans ce chapitre. Le réalisme cognitiviste a trouvé les faveurs de bien des philosophes. Il consiste d'une part à dire qu'il existe des propriétés morales (lesquelles garantissent qu'une assertion morale peut être vraie ou fausse selon qu'elle correspond ou non à la réalité morale)[332], d'autre part à postuler que les gens possèdent les moyens cognitifs nécessaires pour saisir ces propriétés (ils ont un accès épistémique à la réalité morale).

Cette position métaéthique est pourtant assez peu compatible avec l'analyse descriptive présentée au chapitre précédent. Nous avons vu qu'aucune pensée, assertion ou action n'est morale indépendamment du contexte ; pour obtenir le qualificatif de « moral », il faut la justification d'un sujet rationnel et la réalisation de la condition altruiste. De plus le tableau affectif parle plutôt en faveur d'une conception projectiviste, selon laquelle les valeurs que nous prônons ne correspondent pas à des propriétés morales qui appartiennent à la situation évaluée ; nous les projetons sur la situation en question.

En principe, il est logiquement possible de distinguer la réalité morale (niveau ontologique) de la psychologie morale, laquelle a fait l'objet de l'étude descriptive du chapitre précédent. Mais s'il existe une réalité morale, celle-ci est suffisante pour délimiter le champ de la moralité. En conséquence, mes deux critères d'individuation de la moralité sont ou bien redondants, ou bien en conflit avec cette réalité ; et le tableau

[332] Un corollaire de cette position est en principe la portée universelle des assertions morales vraies.

affectif devrait être complété voire même fondamentalement révisé afin d'intégrer la question de l'accès épistémique aux propriétés morales. Ce chapitre a pour objectif de nier la réalité ontologique de la morale et de proposer une position métaéthique antiréaliste compatible avec l'analyse descriptive élaborée au chapitre précédent.

6.1. L'impossibilité du réalisme cognitiviste moral en éthique évolutionniste

6.1.1. Définition du réalisme cognitiviste

Les notions de réalisme moral et de cognitivisme moral ont reçu diverses définitions plus ou moins compatibles entre elles.[333] Il est donc important de préciser ce que j'entends par ces termes. Dans le cadre de cet ouvrage, le « réalisme » désigne la thèse selon laquelle le bien et le mal sont des propriétés morales (appartenant aux états de fait ou aux individus)[334] qui ne dépendent pas des croyances et attitudes individuelles des gens mais possèdent une réalité propre et indépendante[335]. La thèse réaliste a des implications au niveau sémantique : elle implique que les assertions morales peuvent être vraies ou fausses selon qu'elles correspondent ou non à la réalité morale.

Par « cognitivisme », il faut entendre la thèse selon laquelle nous sommes capables, d'une manière ou d'une autre (par exemple par le biais de l'intuition morale ou des émotions), d'avoir un accès cognitif à ces propriétés morales.

[333] A ce propos, voir BLACKBURN 2000/1998, p. 120.

[334] Bien des auteurs (notamment CASEBEER, RUSE, ROTTSCHAEFER) utilisent le terme « valeur morale » pour désigner, soit ce que j'appelle « propriété morale », soit un autre type de réalité métaphysiquement vague. Toutefois, dans le cadre de cet ouvrage, j'ai choisi de réserver la notion de valeur pour désigner les concepts abstraits que nous appliquons aux états de fait en fonction de nos préférences (voir section 5.4.2).

[335] Beaucoup d'auteurs (notamment CASEBEER, RUSE, ROTTSCHAEFER) utilisent le terme « objectivisme » pour désigner ce que j'appelle « réalisme ». La notion d'objectivité étant définie de toute sorte de manières, je retiens le terme « réalisme » afin d'éviter toute confusion.

Ces deux thèses sont généralement défendues conjointement car la seconde implique la première[336] et il paraît absurde d'affirmer l'existence de propriétés auxquelles nous n'aurions aucun accès cognitif. David BRINK par exemple écrit: [337]

> « Un réaliste moral pense que les assertions morales devraient être comprises de manière littérale ; il y a des faits moraux et des propositions morales vraies, et les jugements moraux prétendent rendre compte de ces faits et expriment ces propositions. (…) [Selon un réaliste moral] l'éthique ne traite pas seulement d'états de fait ; elle traite de faits qui existent indépendamment des croyances des gens sur ce qui est bon ou mauvais. » (BRINK 1989, p. 20, ma traduction)[338]

Si le réalisme cognitiviste s'avère une position acceptable, on peut en tirer des conclusions pour l'éthique normative ; en effet, cette position métaéthique porte à croire qu'il est possible de produire des énoncés moraux vrais et justifiés. Ils sont *vrais* s'ils reflètent adéquatement les propriétés morales qui existent dans le monde (ces dernières étant les conditions de vérité objectives et indépendantes de la pensée de ceux qui les prononcent) et *justifiés* si l'on peut faire confiance à notre accès cognitif à ces propriétés.[339]

---

[336] On dit parfois que la théorie de l'erreur (dont il sera question plus loin) est une position à la fois cognitiviste et antiréaliste. Mais dans ce cas, on attribue un autre sens à la notion de cognitivisme. A ce propos, voir la note 388.

[337] David BRINK parle uniquement de réalisme mais sa définition comporte une dimension cognitive et sémantique.

[338] « A moral realist thinks that moral claims should be construed literally; there are moral facts and true moral propositions, and moral judgments purport to state these facts and express these propositions. (…) [According to a moral realist] not only does ethics concern matters of facts; it concerns facts that hold independently of anyone's beliefs about what is right or wrong. » (BRINK 1989, p. 20)

[339] On pourrait reprocher à cette définition du réalisme moral d'être trop contraignante, d'impliquer un trop important engagement ontologique (SAYRE-MCCORD 2005, § 3). Toutefois, il me semble que si l'on émousse l'aspect ontologique de la position réaliste, on ne comprend plus vraiment en quoi elle se distingue des positions antiréalistes, du moins de ses versions modérées comme celle de Gilbert HARMAN (2000) par exemple. Si l'on refuse d'affirmer explicitement l'existence d'entités morales indépendantes de nos attitudes et croyances, premièrement, je ne vois pas en quoi il serait encore utile de parler de « réalisme », deuxièmement, il me semble que le débat se réduirait à la question de savoir si les assertions morales peuvent être considérées comme vraies ou fausses (et en quel sens il faut comprendre cette vérité et cette fausseté) ; mais ce débat relève davantage de la question du *relativisme* moral que de celle du *réalisme* moral.

Il y a plusieurs manières d'être réaliste cognitiviste moral, c'est-à-dire différentes manières de concevoir la nature des faits moraux, ainsi que notre accès cognitif à ces faits. Le schéma ci-dessous représente diverses interprétations des propriétés morales qui seront analysées plus en détail dans cette section : les propriétés morales peuvent être considérées comme naturelles ou non naturelles. On peut penser que les propriétés non naturelles sont de type Idéel, surnaturel ou survenant. Quant aux propriétés naturelles certains auteurs les expliquent en termes darwiniens et d'autres non. Enfin il existe d'autres distinctions comme le caractère *response-dependent* (c'est-à-dire de dépendre de la manière dont les êtres humains réagissent de par leur nature) ou non *response-dependent* des propriétés, ou le fait qu'elles soient dépendantes ou non de la nature humaine. Ces distinctions deviendront plus claires au fil du texte.

<table>
<tr><td rowspan="3">Propriété Morale non naturelle</td><td colspan="2">Idéelle</td><td colspan="2">PLATON</td></tr>
<tr><td colspan="2">Surnaturelle</td><td colspan="2">Théorie du commandement divin</td></tr>
<tr><td colspan="2">Survenante irréductible à des propriétés non morales</td><td colspan="2">MOORE (version propriétés indéfinissables, inanalysables)
SAYRE-MCCORD
MCDOWELL (version response-dependency)[340]</td></tr>
<tr><td rowspan="4">Propriété morale naturelle</td><td rowspan="3">Darwinienne</td><td rowspan="2">Réductible à des propriétés non morales</td><td>Indépendante de la nature humaine</td><td>SPENCER
EIBL-EIBESFELDT
E. WILSON</td></tr>
<tr><td>Dépendante de la nature humaine</td><td>ARNHART
CASEBEER
RICHARDS</td></tr>
<tr><td>Survenante irréductible à des propriétés non morales</td><td colspan="2">ROTTSCHAEFER (version response-dependency)
PRINZ (version response-dependency)</td></tr>
<tr><td rowspan="2">Non darwinienne</td><td>Réductible à des propriétés non morales</td><td colspan="2">Utilitarisme décrit par MOORE (1903)
JACKSON</td></tr>
<tr><td>Survenante irréductible à des propriétés non morales</td><td colspan="2">Réalisme de Cornell</td></tr>
</table>

Nous verrons que parmi les différentes options ouvertes au réalisme cognitiviste moral, certaines ne peuvent pas être adoptées par un tenant de l'éthique évolutionniste,

[340] Notons que MCDOWELL se décrit comme naturaliste mais sa conception du naturalisme ne correspond pas à celle qui est utilisée ici (voir p. 276). Selon cette dernière, MCDOWELL se classe dans le camp des antinaturalistes.

et cela pour des raisons inhérentes à l'approche évolutionnaire. D'autres formes de réalisme cognitiviste font un mauvais usage des théories évolutionnistes. Enfin les approches apparemment les plus crédibles échouent dans leur objectif ou sont en bute à différentes objections, notamment un argument de type évolutionnaire (l'acide universel). En définitive, l'impossibilité d'être réaliste évolutionniste en ressortira.

6.1.2. Le réalisme non naturaliste

Considérons trois formes de réalisme cognitiviste non naturaliste avant de leur adresser à toutes trois une même critique.

La première position, le réalisme platonicien, est généralement considérée comme une forme paradigmatique de réalisme cognitiviste. Le réaliste platonicien défend l'existence d'une Idée de bien. L'Idée de bien ne dépend en rien de ce que pensent les gens à son sujet ; elle n'est ni une abstraction ni une forme de notre entendement que nous appliquons aux choses. D'autre part, l'Idée de bien n'est pas dans les objets; elle est en soi et d'une manière absolue.[341] Les sujets moraux ont un accès cognitif plus ou moins confus à cette Idée par la contemplation intellectuelle (PLATON, *République* VI 505a-b).[342]

La seconde position est le surnaturaliste selon lequel le bien correspond à la volonté d'un être surnaturel, généralement Dieu. Puisque les propriétés morales émanent ou participent d'un être divin, on peut dire qu'elles sont surnaturelles.[343]

---

[341] En réalité, la théorie de PLATON est un peu plus complexe que cela. Il s'agit d'un réalisme asymétrique au sens où il n'y a pas d'Idée de mal, ce dernier étant compris comme un manque de bien. D'autre part, l'Idée de bien entretient un rapport avec le monde sensible : les choses du monde sont plus ou moins bonnes selon qu'elles « participent » plus ou moins à l'Idée de bien. Ainsi, au moyen de cet effet de participation, le bien se trouve également dans le monde.

[342] Plus précisément, selon PLATON, pour avoir accès au monde des Idées, les hommes doivent appliquer la méthode dialectique ; il s'agit de se détacher par la pensée des choses sensibles et de s'élever jusqu'au monde des Idées. Après avoir contemplé les Idées, la pensée redescend dans le monde sensible pour l'exercice concret de l'activité morale et politique (*République*, livre VII).

[343] Notons que toute théorie morale religieuse ne doit pas forcément souscrire au réalisme surnaturel. On peut par exemple imaginer que Dieu a créé des propriétés morales naturelles (au même titre que les autres objets de ce monde) ou que les commandements divins ne sont pas en adéquation avec des propriétés morales (position défendue par les nominalistes comme Guillaume D'OCCAM).

L'accès cognitif à ces propriétés se fait par le biais d'expériences mystiques ou plus simplement via la lecture des textes sacrés.

En troisième position, le réalisme non naturaliste de la survenance défend l'idée que les propriétés morales ne sont ni naturelles (au sens où elles ne peuvent pas être étudiées par les sciences naturelles et sociales), ni surnaturelles ni Idéelles. Elles sont inanalysables et *surviennent* sur les propriétés naturelles. La thèse de la survenance concerne les rapports entre des propriétés différentes d'un seul et même état de choses. La notion de survenance peut être utilisée dans différents contextes mais ici, seul celui des actions morales nous intéresse. Un état de choses (par exemple l'action de battre son petit frère pour le plaisir) possède des propriétés naturelles (physiques, chimiques, …). Selon la thèse de la survenance, certains états de choses, en plus de leurs propriétés naturelles, possèdent également des propriétés morales survenantes (dans notre exemple, la propriété d'être mauvais). Une caractéristique particulière de la notion de survenance utilisée en philosophie morale est que la relation de dépendance entre les propriétés naturelles et survenantes est asymétrique : si deux états de choses ont exactement les mêmes propriétés naturelles, alors elles ont exactement la même propriété morale (à condition évidemment qu'elles aient une pertinence morale), mais l'inverse n'est pas le cas. Deux états de choses peuvent avoir la même valeur morale tout en ayant des propriétés naturelles tout à fait différentes ; par exemple l'acte de vendre des armes à un groupe terroriste ou l'acte de subtiliser de l'argent destiné à une association humanitaire sont tous deux moralement mauvais. L'intérêt de cette asymétrie est que les propriétés morales dépendent des propriétés naturelles sans pour autant pouvoir y être réduites.

Quant à l'accès cognitif à ces propriétés morales non naturelles survenantes, on peut imaginer l'existence d'un sens moral (HUTCHESON 1991/1738) ou d'une intuition (MOORE 1903) ou le recours aux émotions (TAPPOLET 2000)[344]. Il y a également, les défenseurs de la thèse du locuteur compétent (*competent user* ; SAYRE-MCCORD 1997, p. 286 note, p. 291) ; ceux-ci ne se prononcent pas vraiment sur le type d'accès cognitif et se contentent d'affirmer que les individus qui défendent la « meilleure théorie morale » peuvent être considérés comme des juges compétents pour définir les

[344] Christine TAPPOLET, au contraire de HUME, attribue une fonction cognitive aux émotions ; pour elle, les épisodes émotionnels sont en réalité des intuitions. Par exemple, la terreur que l'on ressent devant un chien est une intuition de la propriété d'être effrayant que possède le chien. Selon elle, il en va de même pour les propriétés morales.

propriétés morales (ces dernières étant causalement responsables de la conception que s'en font les locuteurs compétents).

A mon avis, ces approches sont insatisfaisantes dans la mesure où elles reposent sur des notions extrêmement floues. Qu'est-ce qu'un sens moral, une intuition morale, une bonne théorie morale ? En cas de désaccord, comment peut-on décider qui a raison, qui a eu la bonne perception, intuition ou réaction émotionnelle? Comment faut-il comprendre la notion même de propriété morale non naturelle ?

D'autre part, ces formes de réalisme cognitiviste paraîtront douteuses pour tout éthicien évolutionniste car ce dernier s'efforce de développer une conception scientifique du phénomène moral. Or les propriétés Idéelles ou surnaturelles sont très étranges d'un point de vue scientifique et en particulier pour une perspective évolutionnaire. George Edward MOORE (1903) et d'autres font appel à la notion de survenance pour expliquer ces propriétés non naturelles, mais cela ne semble pas les rendre moins étranges. Ici, il vaut la peine de mentionner le fameux argument de l'étrangeté formulé par John MACKIE (1977, pp. 38-42). Dans sa version ontologique, cet argument dit que les propriétés survenantes sont étranges car on ne parvient pas à saisir la relation qui lie la propriété morale qui caractérise une situation et les propriétés naturelles de cette situation. La théorie de la survenance pose une dépendance des propriétés morales par rapport aux propriétés naturelles ; si certaines propriétés naturelles de la situation changent, la propriété morale survenante changera également ou disparaîtra. Mais en même temps, cette même théorie affirme que les conditions de vérité des assertions relatives aux propriétés survenantes ne sont pas les mêmes que celles relatives aux propriétés naturelles sur lesquelles elles surviennent. Ces deux caractéristiques de la survenance mises ensemble engendrent des difficultés logiques et conceptuelles extrêmement difficiles à surmonter. Comment peut-on saisir ce genre de relation lorsque l'on observe une situation moralement pertinente ? Comment distinguer les propriétés survenantes des propriétés sur lesquelles elles surviennent et le lien asymétrique entre les deux ? MACKIE en conclut que la situation serait bien plus simple et compréhensible si l'on se passait de ces propriétés censées survenir sur certaines

caractéristiques naturelles ; il est préférable de les remplacer par de simples réponses subjectives causalement liées à ces caractéristiques naturelles.[345]

Un défenseur du réalisme pourrait s'accommoder de cette étrangeté[346] et pourrait objecter ici que l'évolution de son côté ne peut opérer sans l'existence préalable de la matière et d'entités capables de se répliquer et donc toutes les propriétés n'entrent pas dans le champ d'une perspective évolutionnaire. Dès lors, pourquoi ne pourrait-on pas admettre que l'existence des propriétés morales n'est pas due au processus de l'évolution ?

Trois réponses peuvent être faites à cette objection. Premièrement, si l'on postule que les propriétés morales sont venues à exister indépendamment du processus de l'évolution, alors on pourrait tout aussi bien postuler l'existence d'une multitude d'autres choses, aussi farfelues soient-elles : des déités, des forces occultes, etc. (SOMMERS & ROSENBERG 2003, p. 659). Deuxièmement, si on admet que l'activité morale est un produit de l'évolution et que l'évolution ne se dirige pas vers un but particulier, il s'ensuit que les valeurs et les normes prônées par les hommes auraient pu être tout autres si l'évolution avait pris une autre direction. Cela implique que s'il existait des propriétés morales indépendantes de la contingence humaine, nous pourrions avoir évolué de manière à penser exactement l'inverse, ce qui paraît ridicule (RUSE 1998/1986).[347] Troisièmement, si l'on défend une position évolutionniste, c'est précisément parce que l'on espère pouvoir trouver une explication scientifique de la réalité morale. Dans le cadre d'une telle entreprise, il faut de bonnes raisons scientifiques pour postuler l'existence de quelque chose ; de plus on tentera de réduire au minimum le nombre de ces postulats. Il y de bonnes raisons scientifiques de postuler

---

[345] « How much simpler and more comprehensible the situation would be if we could replace the moral quality with some sort of subjective response which could be causally related to the detection of the natural features on which the supposed quality is said to be consequential [or supervenient]. » (1977, p. 41)

[346] MOORE par exemple ne voit aucune difficulté à affirmer que les propriétés morales non naturelles sont fondamentalement « non analysables ».

[347] « For the Darwinian, what works is what counts. Had evolution taken us down another path, we might well think moral that which we now find horrific, and conversely. This is not a conclusion acceptable to the traditional objectivist [ici, « objectivist » signifie « réaliste moral »]. » (RUSE 1998/1986, p. 254) « What this all means is that, whatever objective morality may truly dictate, we might have evolved in such a way as to miss completely its real essence. (…) Clearly, this possibility reduces objectivity in ethics to a mass of paradox. » (RUSE 1998/1986, p. 108 ; voir aussi 1993)

l'existence de la matière et l'apparition d'entités capables de se répliquer ; toute la théorie de l'évolution repose sur ces deux postulats. Il n'en va pas de même des propriétés morales. Nous avons vu au chapitre 5 comment il est possible de rendre compte de notre activité et pensée morales sans postuler l'existence d'entités aussi étranges que des propriétés Idéelles, surnaturelles, ou survenantes et inanalysables (à ce propos, voir aussi BLACKBURN 1993, p. 180).

En conclusion, nous pouvons affirmer qu'un défenseur d'une approche évolutionnaire de l'éthique ne peut pas adopter une position non naturaliste en métaéthique.

6.1.3. Le réalisme naturaliste non darwinien

Les formes de réalisme cognitiviste non naturalistes ayant été exclues (jusque là, aucun éthicien évolutionniste ne me contredira), voyons ce qu'il en est des positions naturalistes. Il en existe plusieurs que je me propose de considérer avant de montrer qu'aucune n'est acceptable.

Par *naturalisme*[348] j'entends le point de vue selon lequel les propriétés morales sont empiriquement ou du moins scientifiquement accessibles tout comme le sont les propriétés physiques telles que la grandeur, la vitesse, etc. Cette position peut se décliner sous différentes formes et a été défendue à la fois dans le cadre de la philosophie analytique traditionnelle (qui ne tient pas compte des données évolutionnaires) et en *métaéthique évolutionniste*. Cette dernière traite des mêmes questions que la métaéthique classique mais s'en différencie en ce qu'elle s'inspire largement des données provenant de sciences comme la biologie de l'évolution, la neurologie, l'économie expérimentale, la théorie des jeux, l'anthropologie et la psychologie évolutionniste.

Dans cette section, je présenterai deux positions naturalistes non darwiniennes, c'est-à-dire des positions dont les défenseurs pensent que les propriétés morales ne sont

[348] Je ne reprends donc pas à proprement parler la conception que se fait George Edward MOORE du naturalisme dans *Principia Ethica*. Pour lui, le naturalisme est une théorie qui prétend pouvoir donner une définition de termes moraux comme « bien » ou « mal » (par exemple « bien » signifie « être plaisant »). Ainsi MOORE conçoit le naturalisme comme une thèse sémantique alors que la définition que je retiens se situe au niveau épistémique.

pas un produit de l'évolution (ou du moins qu'il n'est pas utile d'en rendre compte de cette manière).

La première est une forme de naturalisme de la survenance qui conçoit les propriétés morales comme des entités irréductibles aux propriétés non morales. La seconde est une version réductionniste.

Un groupe de philosophes de l'Université Cornell (notamment Richard BOYD, Nicholas STURGEON et David BRINK) a développé une position que l'on appelle « réalisme de Cornell ». Ces auteurs s'inspirent largement de la philosophie du langage et de la philosophie des sciences. Selon eux, les gens sont capables d'utiliser des termes comme « bien » ou « mal » et de leur donner une signification qui correspond plus ou moins à une réalité extérieure composée uniquement de faits naturels. Ces derniers instancient des propriétés morales survenantes naturelles.[349] Les réalistes de Cornell ajoutent que l'usage des termes moraux est causalement régulé par les propriétés naturelles en question. Mais cela ne signifie pas pour autant que les termes moraux sont de simples équivalents de prédicats non moraux ; ces derniers ont plutôt pour fonction d'expliciter ou éclairer les termes moraux qui eux sont irréductibles. Concrètement, tout comme pour beaucoup d'autres termes, on peut donner du prédicat « bien » une définition naturelle qui révèle l'essence de la propriété qu'il exprime (Richard BOYD 1988) ; tout comme « eau = H_2O » peut être considéré comme une vérité nécessaire,[350] il est possible, au moyen de longues observations empiriques, de produire des vérités morales nécessaires en découvrant des correspondances entre les concepts moraux que l'on utilise et les propriétés naturelles qu'ils reflètent.[351] Autrement dit, l'analyse de

[349] De par leur caractère survenant, même si elles ne peuvent se révéler qu'au travers de propriétés naturelles, les propriétés morales ne sont pas réductibles à un ensemble donné de propriétés naturelles (car elles sont susceptibles de diverses sortes de réalisation). Pour cette raison, il n'est pas évident d'en donner une définition exacte.

[350] C'est une vérité synthétique et non analytique car le seul moyen de prouver cette définition est de mener des expériences empiriques.

[351] Les réalistes de Cornell établissent une analogie stricte entre le domaine scientifique et le domaine moral : de même que pour les énoncés scientifiques, les énoncés moraux ne peuvent pas être testés empiriquement isolément par rapport à d'autres énoncés car selon eux, toute observation est chargée de théorie ; mais cela ne leur pose pas de problème particulier car à l'image de la connaissance scientifique, la connaissance morale est soit fournie par la cohérence épistémique (BRINK 1989) soit formée dans le cadre du développement d'un système holistique (*confirmational holism* ; Richard BOYD 1988; BRINK

l'usage que font les gens des termes moraux permet au philosophe de détecter progressivement la réalité morale causalement responsable de cet usage.

A première vue, la théorie des penseurs de Cornell est séduisante. Toutefois, elle s'avère bien peu convaincante dès que l'on cherche à dépasser le stade programmatique. Les vérités morales que cette théorie est censée permettre de découvrir ne semblent pas si aisées à découvrir. En effet, si dans bien des domaines, il est possible de recourir à une autorité institutionnalisée pour s'entendre sur la signification des mots et déterminer leurs référents (par exemple le dictionnaire, une communauté scientifique, etc.), il ne semble pas que ce soit le cas pour le domaine moral. Comme le dit Simon BLACKBURN (2000/1998) l'éthique traite de concepts essentiellement contestables si bien qu'il n'est gère possible de recourir à une seule autorité pour trancher en cas de désaccords car cette autorité serait elle-même contestée.[352] Dans le même ordre d'idée, David ZIMMERMAN (1984, pp. 90-91) fait remarquer qu'en éthique, contrairement aux sciences, on ne dispose pas d'espèces naturelles stables ou suffisamment neutres du point de vue théorique (comme l'eau ou l'or) pour pouvoir servir d'objet d'observation. Dans le domaine moral, il est donc utopique d'obtenir des résultats que l'on pourrait caractériser de « scientifiquement valides ». Ainsi, il est difficile d'imaginer que des données observationnelles puissent réellement contribuer à l'acquisition d'une connaissance morale.

Une autre manière de défendre une position réaliste cognitiviste naturaliste inspirée de la philosophie du langage est de réduire purement et simplement les termes moraux à des termes non moraux, cette définition étant ensuite censée refléter la réalité extérieure. Il s'agit d'une forme de réalisme cognitiviste naturaliste *réductionniste* (du moins au niveau sémantique), une forme plus radicale que la position défendue par les réalistes de Cornell. Pour ces derniers, la cognition morale ne se fait pas forcément de manière directe ; tout ce qu'ils affirment, c'est que les instanciations des propriétés morales sont causalement responsables de la manière dont nous formons nos concepts

---

1989). L'idée est que tout énoncé fait partie d'un ensemble ou d'un réseau d'énoncés liés par des relations complexes de dépendance réciproque ; ainsi, il sera considéré comme plus ou moins justifié en fonction de sa contribution à la cohérence du réseau auquel il appartient.

[352] « In ethics we have no single authority, and we are dealing with 'essentially contestable concepts' » (BLACKBURN 2000/1998, p. 200).

moraux. Au contraire, les naturalistes réductionnistes s'avancent davantage et affirment d'une part que nous avons un accès cognitif direct aux propriétés morales, d'autre part, qu'au sein même du langage, il est possible de donner des définitions réductrices de tous les concepts moraux, aussi complexes soient-ils. Selon eux, nos pratiques morales peuvent être décrites en termes purement descriptifs et ces descriptions peuvent être comprises même si l'on ne possède aucun concept moral ; de plus, les deux parties de ces descriptions (ou définitions réductrices, lorsque la description est complète) reflètent, ou plus exactement, sont déterminées par les mêmes propriétés naturelles.[353]

Frank JACKSON (1998) et Philip PETTIT défendent une telle position. Pour eux, les propriétés morales surviennent sur d'autres propriétés naturelles tout en étant elles-mêmes naturelles.[354] Elles sont naturelles (ou au moins analogues aux propriétés naturelles) au sens où on peut en donner des descriptions complètes exemptes de notions normatives. Ces descriptions peuvent être élaborées en analysant les pratiques d'une moralité populaire mature (*mature folk morality*), c'est-à-dire une moralité pratiquée par des individus doués d'esprit critique. Cette réduction descriptive doit être comprise en un sens holistique ; le contenu d'une assertion morale ne prend sens que lorsque l'on détermine simultanément le sens des autres assertions produites dans un contexte complexe ; les termes moraux sont spécifiés par rapport à leur rôle dans la théorie morale populaire.[355]

Cette position semble mener le projet réductionniste plus avant, même si dans les faits, elle ne propose aucune définition concrète du bien et du mal moral. Elle est également plus « palpable » dans la mesure où elle ne se contente pas d'observer la manière dont les gens en général pratiquent la morale ; elle accorde du crédit

---

[353] « It is true that, according to the functionalist theory, the total body of descriptive information entails the evaluative way things are. » (JACKSON & PETTIT 1995 p. 28)

[354] Dans ses écrits plus récents, JACKSON préfère qualifier les propriétés morales de *descriptives* (plutôt que de *naturelles*) parce que nous attribuons ces propriétés aux choses et aux actions au sein du langage descriptif (JACKSON & PETTIT 1996, pp. 82-83).

[355] « The moral terms are specified by their role in received moral theory – in folk moral theory – and while this theory has a purely descriptive content (…) the content of any one claim is fixed only so far as the contents of others are fixed simultaneously. Moral terms are reducible to descriptive terms, at least in principle, but the reduction involved is holistic, not atomistic. » (JACKSON & PETTIT 1995, p. 24) « The meanings of ethical and evaluative terms are given by their place in a complex network. We call this network 'mature folk morality'. » (JACKSON & PETTIT 1996, p. 83)

uniquement aux pratiques morales des individus doués d'esprit critique.[356] Toutefois elle est liée à différents problèmes. JACKSON n'explique par pourquoi les gens doués d'esprit critique sont plus aptes à déceler les propriétés morales que les autres. Il s'agit là d'une affirmation a priori qu'il n'est pas aisé de défendre. De plus, on peut raisonnablement douter de la capacité des gens doués d'esprit critique de s'entendre sur la définition des termes moraux. Il semblerait donc que la position de JACKSON souffre des mêmes faiblesses que le réalisme cognitiviste de Cornell ; elle se trouve en mal d'autorité pour trancher les désaccords au sujet de concepts essentiellement contestables.

L'objection principale que l'on peut faire à ces deux formes de réalisme cognitiviste est qu'elles passent un peu rapidement du niveau sémantique au niveau ontologique. Décrire le bien et le mal moral en analysant la manière dont les gens font usage des termes moraux ne donne aucun droit d'inférer directement ce qui existe réellement. Le passage du niveau sémantique au niveau ontologique ne peut pas se faire de manière aussi directe. Pour garantir un tel passage, il faut postuler l'existence de capacités cognitives qui permettent de saisir cette réalité. Or si l'on prend au sérieux la théorie de l'évolution, il faut admettre que cette capacité cognitive en est un produit. Cette dernière phrase mérite une plus longue explication. L'idée que j'aimerais défendre ici est que si l'on veut être un éthicien évolutionniste conséquent, il faut rejeter l'idée même d'un sens moral, c'est-à-dire d'une capacité qui nous fournit un accès cognitif à des propriétés morales survenantes. Cela est dû à l'application de l'« acide universel ». Je m'explique. L'argument de l'acide universel a été formulé par Daniel DENNETT (1995) et récemment étendu au domaine moral par Tamler SOMMERS et Alex ROSENBERG (2003). DENNETT commence par constater deux aspects déroutants de l'évolution. Premièrement, il s'agit d'un processus entièrement aveugle, non intentionnel, dénué de but ou de finalité. Or, malgré ces apparentes limitations, ce processus est capable de créer un ordre extrêmement complexe et organisé : le monde tel que nous le connaissons. Deuxièmement, la sélection naturelle qui est l'élément clé du processus évolutionnaire peut opérer sur différentes sortes de substrats (les exemples

[356] Selon Stephen YABLO (2000), JACKSON échoue dans son analyse réductrice des termes éthiques car il fait appel de manière a priori à un terme évaluatif : le caractère *mature* de la moralité populaire de référence. A cette critique, JACKSON pourrait rétorquer que la maturité est simplement révélatrice d'esprit critique. Il s'agit alors de se demander si l'attribution du qualificatif « esprit critique » est évaluative.

paradigmatiques étant les gènes et les entités culturelles). En référence au fait que la sélection naturelle est neutre par rapport au substrat et au fait que le darwinisme explique comment la complexité du monde que nous connaissons est le résultat de mécanismes dépourvus de dessein, DENNETT qualifie joliment la théorie darwinienne d'« acide universel » qui transforme tous les domaines théoriques sur son passage.

> « L'acide universel est un liquide si corrosif qu'il mangera n'importe quoi ! (…) Il dissout les bouteilles de verre et les cannettes d'acier aussi facilement que des sacs en papier. Que se passerait-il si vous rencontriez ou créiez une cuillerée d'acide universel ? La planète tout entière serait-elle finalement détruite ? Que laisserait-elle dans son sillage ? Après que tout aurait été transformé par cette rencontre avec l'acide universel, à quoi le monde ressemblerait-il ? (…) [La ressemblance de l'idée de Darwin avec l'acide universel est frappante : l'idée de Darwin] dévore absolument tous les concepts traditionnels et laisse dans son sillage une vision du monde révolutionnée ; où presque tous les anciens jalons sont toujours reconnaissables, mais transformés de fond en comble. » (DENNETT 2000, p. 71)

Si l'on accepte la théorie de l'évolution comme une manière éclairante de comprendre le monde et que l'on tire toutes les conséquences de cette théorie, il faut admettre que tous les systèmes biologiques (y compris les êtres humains) et tout ce qui les caractérise peut être expliqué comme un résultat de l'évolution. Si l'on admet que les êtres humains possèdent la capacité de saisir les propriétés morales, alors l'acide universel peut être appliqué au niveau de la réflexion métaéthique ; cette capacité doit être conçue comme le fruit d'une pression sélective exercée sur les êtres humains (c'est le cas même s'il s'agit d'un produit dérivé d'autres adaptations). Citons ici DENNETT :[357]

> « L'éthique est-elle un domaine entièrement autonome d'enquête ? Flotte-t-elle, totalement détachée de faits relevant d'une quelconque autre discipline ou d'une quelconque tradition ? Nos intuitions morales proviennent-elles d'un quelconque module éthique implanté dans nos cerveaux (ou dans nos 'cœurs', pour employer le

[357] En réalité, la morale est le seul domaine auquel DENNETT applique l'acide universel de manière trop parcimonieuse ; même s'il accepte que l'acide discrédite l'idée de sens moral, il ne se résout pas à accepter l'absence de réalité ontologique des propriétés morales. SOMMERS et ROSENBERG (2003) lui reprochent à juste titre ce manque de cohérence.

> langage traditionnel) ? Ce serait un crochet céleste trop douteux pour qu'on puisse y accrocher nos convictions les plus profondes quant à ce qui est bien ou mal » (DENNETT 2000, p. 538).

Est-il possible de produire une explication évolutionnaire du sens moral, c'est-à-dire de la capacité de saisir le bien ou le mal que présentent certains états du monde ? DENNETT a raison d'en souligner la difficulté. Une solution serait de dire que le sens moral est un produit dérivé d'autres adaptations. Toutefois, cette approche est peu convaincante car si le sens moral existe, il doit s'agir d'une capacité à la fois raffinée et fondamentale (de manière similaire à d'autres sens comme la vue ou l'ouïe). Or il est difficile d'imaginer qu'une telle capacité soit un simple effet dérivé et surtout de quelle autre adaptation elle serait dérivée. La solution la plus probable serait de dire qu'elle a été sélectionnée parce qu'elle apportait un avantage sélectif.[358]

Il est assez aisé d'imaginer que la capacité de concevoir des normes morales, poser des jugements moraux et agir en fonction de ces normes ou jugements sont apparues et se sont maintenues au fil de l'évolution pour répondre à un besoin ; celui de renforcer les contacts sociaux. Les hommes qui possédaient ces capacités ont été en mesure de former de grandes communautés coopératives. Les nombreux avantages qu'ils ont pu tirer de ce mode de vie leur ont permis de transmettre leurs capacités aux générations suivantes. On peut également reconstruire l'évolution de mécanismes biologiques comme les sentiments empathiques qui nous incitent à poser certains jugements plutôt que d'autres. Mais tout cela ne nous dit rien sur une prétendue capacité de saisir des réalités morales extérieures pas plus que sur la vérité de nos normes et jugements moraux. Comme le remarquent SOMMERS et ROSENBERG (2003, p. 660), la capacité de produire des normes et jugements moraux et la capacité de détecter les faits moraux sont deux phénomènes distincts qui ne sont pas forcément liés. Il est aisé de trouver une explication évolutionnaire pour la première mais non pour la seconde. Cette dernière repose sur deux postulats : l'existence d'une réalité morale extérieure et le caractère adapté (au sens évolutionnaire) de la connaissance de la vérité morale.

Le premier postulat est évidemment problématique car on ne sait pas vraiment en quoi consisterait cette réalité morale. Au fond les propriétés morales survenantes

[358] Et en principe, on admet également l'avantage sélectif du sens moral est toujours opérant, c'est-à-dire que le sens moral n'est pas un simple vestige qui s'est avéré adapté mais a perdu cette propriété au fil de l'évolution.

invoquées par les réalistes de Cornell ou par JACKSON, même si elles sont qualifiées de « naturelles » par ces auteurs, sont ontologiquement à peu près aussi floues que les propriétés non naturelles dont il était question à la section précédente et la version ontologique de l'argument de MACKIE présentée plus haut (p. 274) s'applique également ici. De plus, même à supposer que l'on fasse une faveur au réalisme en lui concédant la réalité de propriétés morales, le second postulat (celui de l'adaptabilité du sens moral) pose problème.

Si l'on raisonne en termes évolutionnaires, on comprend que l'aspect crucial pour l'évolution d'une capacité telle que le sens moral est qu'elle permette aux êtres humains d'agir de manière adaptative sur le long terme. Puisque l'homme est un être social, son comportement adaptatif est partiellement composé d'actes coopératifs, altruistes et d'entraide ; et c'est précisément ce genre d'actions dont on pense que les sens moral est responsable.[359] Mais pour que la diffusion de ce genre d'actes soit garantie, il n'y a aucun besoin de doter l'homme de connaissances au sujet d'une prétendue vérité morale. Il suffit qu'il soit motivé à réaliser ces actions et convaincu de leur bien-fondé. Il est difficile d'imaginer qu'à cet effet, il ait été doté d'une faculté sans aucun doute coûteuse du point de vue de l'évolution si le même résultat peut être produit au moyen de biais psychologiques et mécanismes plus simples.

En définitive, l'application de l'acide universel montre qu'un éthicien évolutionniste ne peut pas défendre une position naturaliste non darwinienne et que s'il défend une théorie qui repose sur la notion de propriété morale survenante, elle n'est pas crédible car aucune explication évolutionnaire convaincante de l'accès cognitif à ces propriétés morales ne peut être fournie.

6.1.4. Le réalisme naturaliste darwinien

Face à l'échec des positions non darwiniennes, il se pourrait que les approches darwiniennes soient plus convaincantes dans leur analyse de ce qu'est le bien et le mal. Si c'est le cas, elles seront peut-être à même d'éclairer la fonction de la capacité

[359] Notons que ce lien ne peut se faire que si l'on ajoute un postulat à mon avis douteux : notre perception du bien et du mal nous motive à agir. Je ne reviendrai toutefois pas ici sur ce débat (à ce propos, voir section 5.2.2).

d'accéder à ces propriétés. Les approches darwiniennes réalistes reprennent le postulat de la réalité des propriétés morales et précisent la nature même de ces propriétés ; non contentes de considérer ce que pensent les gens, elles fournissent des définitions scientifiques des propriétés morales. Ces définitions sont censées à la fois décrire complètement les termes moraux et correspondre à la réalité ontologique ; cette correspondance parfaite entre les niveaux sémantique et ontologique revient à admettre que les propriétés morales sont réductibles à des propriétés non morales.

Il n'est pas forcément évident d'interpréter en termes métaéthiques la pensée d'un bon nombre d'auteurs évolutionnistes car ils ne se sont pas toujours posé la question de l'existence de propriétés morales et de notre accès cognitif à ces propriétés (notamment, SPENCER, EIBL-EIBESFELDT et E. WILSON). Il y aura donc une part d'interprétation dans ce qui suit. On peut toutefois noter une tendance marquée, chez un bon nombre de penseurs évolutionnistes, à soutenir une position naturalisme fortement réductionniste. Ils accordent une grande place à la théorie darwinienne pour expliquer le phénomène moral et mènent jusqu'au bout le projet d'une définition du bien moral ; cela les mène non seulement à un réductionnisme descriptif typique de la position de JACKSON mais également à un réductionnisme ontologique. Considérons quelques exemples.

Plus que quiconque, c'est Herbert SPENCER, un contemporain de DARWIN, qui a tenté d'établir un lien entre l'évolution et l'éthique (1981/1893 ; 1879). SPENCER était un grand optimiste. Selon lui, dans la nature, on peut observer partout du progrès. La notion de progrès, SPENCER la comprenait dans le sens où tous les produits de la nature évoluent vers des formes plus complexes (l'organisme le plus complexe de tous étant bien entendu l'être humain). De plus, il considérait que ce qui est plus complexe (c'est-à-dire le plus évolué) est de plus grande valeur. Ainsi l'évolution, cette force progressiste, par sa nature, nous imposerait l'obligation morale de soutenir le processus de l'évolution qui se dirige vers des formes plus nobles. En d'autres termes, la propriété morale pourrait être réduite à celle de progrès, c'est-à-dire à la plus grande complexité évolutionnaire.

Dans la même veine que SPENCER, l'ethnologue Irenäus EIBL-EIBESFELDT écrit :

> « Tous les organismes qui vivent aujourd'hui sont redevables du fait que leurs parents et leurs ancêtres se sont comportés de manière à pouvoir se reproduire. Ils se sont

> comportés de manière adaptative et par là correctement. » (EIBL-EIBESFELDT 1990/1988, p. 182, ma traduction)[360]

En outre, tout comme SPENCER, EIBL-EIBESFELDT défend une vision progressiviste du processus évolutionnaire, c'est-à-dire qu'il accorde une valeur normative au progrès. Ainsi, en écrivant que nos ancêtres ont pu transmettre leur matériel génétique grâce au fait qu'ils se sont comportés de manière adaptative et par là « correctement », il exprime l'idée qu'un comportement adaptatif est bon parce qu'il manifeste d'un progrès évolutionnaire.

Comme dernier exemple, voici un passage de Edward O. WILSON :

> « L'évolution culturelle des hautes valeurs morales peut-elle avoir une direction et un mouvement intrinsèques et se substituer ainsi totalement à l'évolution génétique ? Je ne le pense pas. Les gènes tiennent la culture en laisse. Cette laisse est très longue, mais inévitablement les valeurs morales subiront des contraintes suivant les effets qu'elles auront sur le patrimoine génétique humain. L'encéphale est un produit de l'évolution. Le comportement humain – comme les aptitudes les plus profondes aux réponses émotionnelles qui le provoquent et le guident – est une façon détournée d'assurer la permanence du matériel génétique humain. La morale n'a aucune autre fonction démontrable. » (E. WILSON 1979/1978, p. 242-243)

E. WILSON prétend que l'éthique peut être abordée uniquement par le bais des sciences naturelles (du moins pour un certain temps). Dans le texte reporté ci-dessus, l'auteur sous-entend que les êtres humains cherchent (même si c'est de manière largement inconsciente) à maximiser leur survie et leur reproduction ainsi que celle de leurs proches et par là, transmettent leur matériel génétique à la génération suivante. Or, en agissant de cette manière ils contribuent à maintenir intact le matériel génétique humain.[361] Abordant la morale dans une perspective purement fonctionnelle, E. WILSON en déduit qu'elle a pour unique fonction de maintenir intact le matériel génétique. Sur

[360] « Alle Organismen, die heute leben, verdanken dies der Tatsache, dass ihre Eltern und deren Vorfahren sich so verhielten, dass sie ihr Erbgut weitergaben. Sie verhielten sich angepasst und damit wohl richtig. » (EIBL-EIBESFELDT, 1990/1988, p. 182)

[361] En d'autres termes, le comportement humain est un mécanisme complexe au moyen duquel le matériel génétique peut rester intact au fil du temps.

ces considérations, il pense pouvoir tirer des conclusions normatives ; c'est ainsi qu'il en vient à affirmer la « valeur fondamentale de la survie des gènes humains » (1979/1978, p. 279) et plus particulièrement du « lot global des gènes humains ».[362] En d'autres termes, E. WILSON semble définir la propriété morale par excellence comme le fait de contribuer au maintien du matériel génétique humain. Cette valeur servira ensuite à soutenir différentes causes dont les droits de l'homme qui garantissent survie et diversité biologique humaine (1979/1978, p. 282).

Je ne pense pas que ces formes de naturalisme réductionniste soient viables. Les positions de SPENCER et EIBL-EIBESFELDT doivent être rejetées d'emblée car elles commettent une erreur grave de méthodologie : elles présupposent que tous les comportements observables sont optimaux avant de chercher une explication en termes d'adaptations antérieures. Autrement dit, elles rationalisent les pratiques actuelles au moyen d'une analyse biologique conçue en termes de progrès (voir ROSENBERG 1991, p. 92 ; voir aussi section 1.1.3, p. 27).

De plus les trois auteurs postulent sans plus d'arguments que le progrès ou le maintien du pool génétique dans le cadre du processus évolutionnaire est *moralement bon*. Or, de manière assez ironique, il semblerait que ces postulats succombent à l'acide universel. En effet, il semblerait que la moralité est spécifique aux êtres humains ou du moins aux organismes dotés de capacités cognitives hautement développées. C'est d'ailleurs un point ces auteurs semblent admettre puisqu'ils appliquent leur théorie aux êtres humains. Dans ces conditions, l'acide universel impose que l'on rende compte de l'évolution de la moralité de manière à mettre en évidence le lien qu'elle entretient avec des capacités particulières aux êtres humains ou à des organismes similaires. Or les comptes rendus réductionnistes de nos trois auteurs indiquent que les propriétés morales n'entretiennent aucun rapport particulier avec la constitution humaine.

Il existe des approches naturalistes darwiniennes qui tiennent compte de la nature humaine. En voici deux exemples paradigmatiques. Larry ARNHART (1998, p. 17), défend l'idée que le bien équivaut à ce qui est désirable du point de vue de la nature

[362] Plus tard, Edward WILSON modifiera substantiellement sa position et se ralliera à celle de Michael RUSE selon laquelle aucune justification morale n'est possible (RUSE & E. WILSON 1989).

humaine, c'est-à-dire à ce qui a été généralement désiré par les êtres humains durant leur histoire évolutionnaire.

> « Si le bon est désirable, alors l'éthique humaine est naturelle dans la mesure où elle satisfait les désirs humains naturels. Il y a au moins vingt désirs naturels qui se sont manifestés de différentes manières dans toutes les sociétés humaines au cours de l'histoire : une vie complète, les soins parentaux, l'identité sexuelle, les relations sexuelles, les liens familiaux, l'amitié, la hiérarchie sociale, la justice comme réciprocité, la régulation politique, la guerre [etc.] » (ARNHART 1998, p. 29, ma traduction)[363]

ARNHART propose une liste de vingt désirs généralement présents chez les êtres humains : par exemple atteindre un statut social élevé, être opulent ou obtenir la justice (au sens de réciprocité). Ainsi, il existe vingt catégories de propriétés morales, correspondant à ces vingt désirs ; et plus généralement, la propriété du bien moral correspond à celle d'être désirable. La moralité découle donc directement de la nature humaine, plus précisément des sentiments et désirs proprement humains.[364]

William CASEBEER, un autre défenseur du naturalisme réductionniste, pense que les propriétés morales sont des relations fonctionnelles propres aux êtres humains. A l'image des réalistes de Cornell, CASEBEER établit une analogie stricte entre le domaine

---

[363] « If the good is desirable, then human ethics is natural insofar as it satisfies natural human desires. There are at least twenty natural desires that are manifested in diverse ways in all human societies throughout history: a complete life, parental care, sexual identity, sexual mating, familial bonding, friendship, social ranking, justice as reciprocity, political rule, war, health, beauty, wealth, speech, practical habituation, practical reasoning, practical arts, aesthetic pleasure, religious understanding, and intellectual understanding. » (ARNHART 1998, p. 29)

[364] En cela, ARNHART s'inspire largement des écrits de HUME et de Robert MCSHEA (1978 ; voir aussi MCSHEA & MCSHEA 1999).

« Parenthood is a human value because human beings have a strong feeling for parental caregiving. Friendship is a human value because human beings have a strong feeling for their friends. Courage in war is a human value because human beings have a strong feeling for patriotic loyalty. Such values are natural to human beings, because such feelings arise from what Hume called 'the original fabric and formation of the human mind, which is naturally adapted to receive them' (1902, [*Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, Oxford: Clarendon Press] p. 172). Pure reason cannot create values because it cannot create feelings. » (ARNHART 1998, p. 80)

scientifique et le domaine moral. Pour lui, la moralité n'est pas inventée mais peut être découverte au moyen de méthodes scientifiques (2003, p. 38).

> « De même que l'on peut avoir une connaissance médicale, on peut avoir une connaissance morale ; cette dernière sera acquise de la même manière que la connaissance scientifique – en raisonnant sur les données de l'expérience. » (CASEBEER 2003, p. 48, ma traduction)[365]

En outre, il mène jusqu'au bout le projet réductionniste en affirmant que les propriétés morales sont des relations fonctionnelles d'un certain type.

> « Si les valeurs [propriétés morales] sont des relations fonctionnelles, et si l'on peut donner une explication 'non étrange' de ce que sont les fonctions (ce qui est certainement possible – à nouveau, pensez à la connaissance médicale), alors les valeurs ne seront pas ces 'entités étranges'[366] indépendantes des faits naturels. Ce seront des entités parfaitement naturelles, que l'on peut déceler et auxquelles on peut attribuer une force explicative au moyen d'une ontologie matérialiste. » (CASEBEER 2003, p. 48, ma traduction)[367]

Puisque ce sont des relations fonctionnelles, les propriétés morales se trouvent à l'interface entre l'organisme et l'environnement ; plus précisément, leur domaine d'extension est ouvert et inclut tous les ensembles de propriétés qui réalisent au mieux la dimension sociale de la nature humaine. En fait, CASEBEER affirme que le moralement louable concerne l'épanouissement du potentiel humain ; mais à y regarder de plus près, on constate qu'il traduit l'épanouissement propre à l'homme par le fait de

[365] « Just as we can come to have medical knowledge, we can come to have moral knowledge; this knowledge will be gained in much the same way that scientific knowledge is – through the application of reason to experience. » (CASEBEER 2003, p. 48)

[366] CASEBEER se réfère ici à un argument de John MACKIE (1977, p. 41) contre l'existence des propriétés morales : si les propriétés morales existaient, alors leur nature ontologique serait bien étrange.

[367] « If values are functional relations, and if we can give a 'non-queer' account of what functions are (this certainly seems possible – again think of medical knowledge), then values will not be theses 'strange entities' that can't be related to natural facts. They will be perfectly natural entities, tractable within and given explanatory force by materialist ontology. » (CASEBEER 2003, p. 48)

naviguer avec succès en société.[368] Ainsi les personnes vertueuses sont celles qui mènent une vie sociale efficace, c'est-à-dire qui réalisent leur fonction d'être social (ce qui est le propre de l'homme) ; et « la connaissance morale correspond à la connaissance de la structure de notre environnement social et de la manière dont on peut y naviguer efficacement. » (CASEBEER 2003, p. 105, ma traduction)[369] En d'autres termes, le bien moral coïncide avec l'efficacité sociale. Quant à l'acquisition de la connaissance morale, elle procède par un apprentissage pratique au moyen d'exemples.

La position de CASEBEER est similaire à celle de ARNHART à la différence que le second n'affirme pas que la moralité consiste en l'épanouissement de la nature humaine. Pour ARNHART les propriétés morales sont directement dépendantes des désirs fondamentalement humains. Par exemple, l'expression de l'amour parental est une propriété moralement bonne car les êtres humains recherchent l'amour parental.[370] En revanche, ARNHART admet que ce qui est désirable pour les êtres humains promeut l'épanouissement de leur humanité.[371] Ainsi, concrètement, en agissant moralement (c'est-à-dire en satisfaisant leurs désirs fondamentaux), nous nous épanouissons en tant qu'être humain, nous réalisons notre vraie nature. Or CASEBEER affirme précisément cela : les agents agissent moralement en réalisant leur nature. Par contre, les vues des deux auteurs divergent lorsqu'il s'agit de définir ce qui caractérise la nature humaine : CASEBEER opte pour la socialité alors que ARNHART relève une vingtaine de désirs qu'il considère comme universels dans l'espèce.

Ces théories sont problématiques pour différentes raisons. Premièrement, elles excluent du domaine moral les actes de sacrifice personnel au profit d'autrui (du moins ceux qui ne correspondent à aucun des désirs fondamentaux ou n'ont pas d'effet

[368] CASEBEER reprend en fait la position de Paul CHURCHLAND (1998) pour qui l'éthique est l'étude de la manière dont les gens doivent se comporter pour s'épanouir, pour actualiser leur potentiel humain ; et selon CHURCHLAND, cela revient à gérer avec succès une vie en société.

[369] « Moral knowledge (…) [is] knowledge of the structure of our social environment and how to navigate effectively within it. » (CASEBEER 2003, p. 105)

[370] C'est pourquoi on peut objecter à ARNHART de confondre ce qui est désiré avec ce qui est désirable, ou du moins de lier causalement l'un à l'autre sans autre explication.

[371] « What is 'desirable' for human beings is whatever promotes their human flourishing. » (ARNHART 1998, p. 82).

bénéfique pour la « navigation efficace » dans l'environnement social), ce qui est assez gênant pour une théorie morale.

Deuxièmement, du fait qu'elles sont centrées sur les individus et leur relation à la nature humaine, ces théories peinent à traiter les questions morales d'ordre général (par exemple le problème de la répartition des richesses ou de l'avortement).

Troisièmement, de par leur caractère universaliste, elles se discréditent mutuellement puisqu'elles ne s'entendent pas sur ce qui compose l'ontologie de la moralité. Au fond, affirmer que les désirs fondamentaux ou la réalisation de la nature humaine caractérisent la moralité n'est rien de plus que l'expression de convictions personnelles des auteurs. Même s'il est évident que l'assouvissement de nos désirs fondamentaux ou la réalisation de notre nature humaine (qu'elle soit réduite ou non à la socialité) sont des aspects fondamentaux de notre vie, rien ne prouve que l'un ou l'autre de ces phénomènes corresponde à une quelconque réalité morale ; ce n'est qu'une affaire de postulats. On pourrait d'ailleurs imaginer une multitude d'autres solutions ; Robert RICHARDS (1986 ; 1993) par exemple, définit la morale par l'altruisme (qu'il comprend au sens de promotion du bien de la communauté)[372]…

En plus de ces diverses critiques, il y a une objection encore plus fondamentale que l'on peut adresser aux théories naturalistes réductionnistes. Il s'agit de l'argument de l'impossibilité de définir le bien moral de manière réductionniste.

George Edward MOORE (1903) a formulé cette idée au moyen de son fameux argument de la question ouverte.[373] L'argument de Moore se situe au niveau sémantique. Nous avons vu que les auteurs naturalistes réductionnistes (SPENCER, ARNHART, etc.) proposent de manière plus ou moins implicite des définitions du bien moral ; et les définitions qu'ils proposent sont censées refléter la réalité, c'est-à-dire que les propriétés morales auxquelles réfèrent les définitions se résument à des propriétés naturelles.

---

[372] RICHARDS ne s'exprime pas sur les questions de métaéthique. En revanche, il tente de déduire des normes à partir de considérations factuelles, définit la morale en termes d'altruisme et pense que nos assertions morales peuvent être vraies ou fausses (voir sections 7.3.1 et 7.3.2). De cela, on peut inférer qu'il souscrirait probablement à une position réaliste.

[373] Dans *Principa Ethica*, MOORE dirige ses objections contre SPENCER (voir p. 284) et John Stuart MILL qui identifie le bien moral à la maximisation de l'utilité (1998/1861); mais l'argument vaut pour toutes les tentatives de définition du bien de manière complète.

Mais d'après MOORE, il n'est pas possible de poser une identité entre le prédicat « bien » et un terme naturel (ou un complexe de termes naturels). Car si deux termes sont synonymes et qu'un locuteur maîtrise les deux termes, il ne peut pas raisonnablement mettre en doute que leur sens soit le même. Or, dans le cas des définitions du bien, il est toujours possible de se demander si les termes naturels censés définir le bien correspondent vraiment au bien. Par exemple, si on reprend la définition de E. WILSON en disant que « bien » signifie « contribuer au maintien du matériel génétique humain », on peut toujours se poser la question « Mais est-ce que le fait de contribuer au maintien du matériel génétique est vraiment bien ? ». Cette question fait sens ; il n'est pas ridicule de se la poser.

Il y a un moyen de proposer une définition du bien et du mal tout en échappant à la critique de MOORE. Cette dernière repose sur une conception très stricte de la définition : selon MOORE, une définition est censée poser une identité logique entre deux termes, c'est-à-dire que les deux termes doivent être synonymes. Il s'ensuit que seules les définitions analytiques (par exemple « un célibataire est une personne non mariée ») peuvent prétendre porter le nom de « définition ». Or ce n'est pas la manière dont les gens font usage des définitions. Par exemple, lorsque l'on dit que l'« eau » signifie « H_2O »,[374] on accepte qu'il s'agit d'une définition même si on ne maîtrise pas complètement le terme « H_2O » et que l'on pourrait se poser la question ouverte « Est-ce que H_2O est de l'eau ? » (PUTNAM 1981, pp. 205-207). Dans cet exemple, on utilise deux termes différents pour référer à la même propriété ontologique. Or en assouplissant les exigences de la définition, c'est-à-dire en faisant fi de la condition de l'identité conceptuelle, on donne à nouveau la place à une définition du bien.[375]

Cela dit, même si l'on rejette l'argument de la question ouverte pour des questions de « définition de la définition », je pense que MOORE a touché un point sensible et que l'on peut reprendre son scepticisme face aux projets réductionnistes au moyen d'autres arguments.

---

[374] Il s'agit ici d'une définition synthétique (car le seul moyen de prouver cette définition est de mener des expériences empiriques) mais néanmoins nécessaire (car elle est scientifiquement prouvée).

[375] Les défenseurs du réalisme de Cornell par exemple échappent à l'argument de MOORE puisqu'ils développent un naturalisme synthétique (par opposition à analytique) ; ils refusent le critère de la synonymie de la propriété d'identité et recherchent une définition synthétique du bien moral (à ce propos, voir STURGEON 1984).

Le premier argument se situe au niveau sémantique et repose sur une considération d'ordre psychologique. L'affirmation selon laquelle tout énoncé moral peut être décrit en termes descriptifs (c'est le cas si l'on peut donner une définition du bien moral) ne permet pas de rendre compte d'un sentiment profondément ancré en nous, selon lequel les énoncés moraux ne sont pas de la même catégorie que les énoncés descriptifs. Or, si tout ce qui est moral peut être décrit en termes descriptifs, il n'y a plus moyen de donner une explication des différences et des rapports qu'entretiennent le moral et le descriptif (par exemple le fait que les termes moraux, au contraire des termes descriptifs, sont prescriptifs par nature).

Le second argument met en question la plausibilité du projet réaliste réductionniste. Il semblerait que si l'on veut mener le projet réductionniste jusqu'au bout en proposant une description du bien moral à la fois claire et exempte de toute composante normative, on perd du même coup l'intérêt de parler de morale. En quelque sorte, cela revient à jeter le bébé avec l'eau du bain. En effet, un projet qui se veut à la fois réaliste et réductionniste doit admettre qu'au niveau ontologique, il y a des réalités morales correspondantes à nos descriptions, c'est-à-dire que les propriétés morales se réduisent à des propriétés factuelles. CASEBEER écrit contre MACKIE que les propriétés morales ne sont pas étranges, qu'elles sont de la même nature que d'autres propriétés naturelles (comme la santé par exemple) et qu'il n'est pas nécessaire de postuler un sens moral pour les saisir.[376] Au fond, pour lui, il suffit de comprendre quelle est l'action socialement la plus avantageuse pour saisir ce qu'il est moralement bon de faire.[377] Mais alors on peut légitimement se demander pourquoi il est encore utile de parler de réalité morale ! A trop vouloir démystifier la morale, on finit par la perdre. Par exemple, si le bien est complètement réduit au socialement efficace, il semblerait que la composante morale recherchée par le réaliste disparaisse tout simplement ; le bien devient une propriété redondante. En bref, il me semble que si l'on veut défendre une position

[376] « The standards for 'health' may vary across organisms, but (contra Mackie) that does not mean that the standards are subjective or that talk about them is laden with error. Value properties are not queer in either the epistemological sense or the metaphysical sense. They are scientifically tractable in the same way that biological notions of function are, and to gain moral knowledge we need posit no 'special sense' above and beyond the traditional tools and methods of scientific naturalism. » (CASEBEER 2003, p. 55)

[377] Il en va de même pour ARNHART qui définit le bien moral en référence à la satisfaction d'un ensemble de désirs.

réaliste morale, il faut précisément éviter une telle réduction sémantico-ontologique qui n'est autre qu'une forme d'éliminativisme.[378]

En résumé, il semblerait qu'un projet réductionniste pleinement assumé revient à accepter que la réalité morale n'est qu'une chimère qui disparait au profit d'une réalité factuelle dès qu'on l'observe de plus près. Dès lors, si l'on tient réellement au réalisme, il faut admettre que la réalité morale survient de manière irréductible sur le factuel, c'est-à-dire ne peut pas être réduite au factuel. Or nous avons vu à la section précédente à quel point ce projet est fragile…

6.1.5. La théorie de la response-dependency

Compte tenu du manque de crédibilité des positions considérées jusqu'à ce point, on peut se tourner vers une autre manière de concevoir la réalité des faits moraux qui a récemment séduit beaucoup de philosophes (MCDOWELL 1985, TAPPOLET 2000 ; ROTTSCHAEFER 1998/1997 ; etc.). Selon la théorie de la *response-dependency*, certains états de fait possèdent des propriétés qui déclenchent des réactions typiques chez certaines espèces d'individus en vertu de la manière dont ces derniers sont constitués. Dans ce contexte, on parle de propriétés *response-dependent* (dépendantes de la réponse des espèces considérées). Au fond les propriétés morales seraient analogues aux qualités secondes de LOCKE, comme le rouge ou le moelleux. Par contre, en plus d'être survenantes à la manière des qualités secondes à la LOCKE, elles seraient non réductibles à des propriétés non morales (que ces dernières soient ou non survenantes ou en relation avec la nature humaine).

L'idée de la *response-dependency* a émergé dans le cadre de la métaéthique traditionnelle ; son défenseur emblématique est John MCDOWELL (1985). Ce qui est particulièrement intéressant pour notre propos est la manière dont cette idée a été reprise dans le cadre de théories métaéthiques évolutionnistes. Il vaut la peine de se demander

[378] ARNHART ne prend pas au sérieux l'argument de MOORE, car selon lui, si on ne pouvait pas passer du factuel au normatif, nous n'aurions aucune raison d'obéir aux normes morales car les seules raisons qui nous poussent à l'action sont d'ordre factuel : nos sentiments et nos désirs (1998, p. 82-83). ARNHART a raison de pointer sur le fait qu'une description du bien n'est pas suffisante pour pousser les gens à agir. Mais au vu de l'argument que je viens de présenter, on peut se demander si son projet réductionniste permet encore de parler de réalité morale...

si les développements théoriques effectués dans ce domaine nous permettront d'accorder plus de crédit au réalisme moral.

Quoiqu'il ne fasse pas usage du terme *response-dependency*, William ROTTSCHAEFER (1998/1997) est un avocat de cette approche. Pour lui, les propriétés morales sont relationnelles : les deux termes de la relation sont l'environnement naturel et social d'une part et le sujet humain d'autre part. Dit autrement, les propriétés morales sont des propriétés à la fois de l'environnement naturel et social auquel un individu est adapté et de l'individu lui-même.[379]

> « Les propriétés morales sont des propriétés biologiques relationnelles biologiquement émergentes. Elles reposent d'une part sur des états de fait [l'environnement naturel et social] d'autre part sur des sujets humains, leurs sentiments, croyances, actions et pratiques. Ces propriétés surviennent sur d'autres propriétés naturelles des personnes et des choses. » (p. 161) « Les deux termes de la relation sont nécessaires et les deux termes obtiennent le label moral en vertu de cette relation. Un terme de cette relation est le sujet humain et ses sentiments et actions morales. L'autre terme est l'environnement naturel et social. » (p. 160) « Les propriétés morales (…) sont réalisées au moyen d'une activité consciente et libre de la part du sujet » (p. 163) (ROTTSCHAEFER & MARTINSEN 1990, ma traduction)[380]
>
> « x peut seulement être bon pour quelque chose ; x ne peut pas simplement être bon. Cela ne signifie pas qu'il n'y a pas de valeurs intrinsèques ; cela signifie simplement que si quelque chose est intrinsèquement valable, cela provient de la relation

[379] Plus récemment, ROTTSCHAEFER a légèrement modifié sa position mais pas forcément dans une direction plus éclairante: les propriétés morales seraient des relations *triadiques* entre les agents, les objets de leur activité et les interactions causales qui lient les deux premiers termes. « [Moral phenomena] are relational natural phenomena involving agents and the objects of their activities, as well as the complex causal interactions that bring about and constitute theses activities. » (ROTTSCHAEFER 1998/1997, p. 222)

[380] « Moral properties are biologically emergent relational properties of things and states of affairs, on the one hand, and human subjects, their moral sentiments, beliefs, actions and practices, on the other. These properties supervene on the other natural properties of persons and things. » (p. 161) « Both terms of the relation are necessary and both terms can be denominated morally because of the relation. One term of that relation is, indeed, the human subject and his or her moral sentiments and actions. The other is the natural and social environment. » (p. 160) « Moral properties (…) are achieved by the conscious, free activity of the person. » (p. 163) (ROTTSCHAEFER & MARTINSEN 1990)

> particulière que cette chose entretient avec une certaine sorte d'agent. » (ROTTSCHAEFER 1998/1997, pp. 219-220, ma traduction)[381]

Selon ROTTSCHAEFER, beaucoup de choses peuvent être comptées comme des biens moraux ; il peut s'agir de propriétés biologiques mais également de propriétés culturelles. Le bien biologique par excellence est la fitness positive. Les biens culturels peuvent être la science, les mathématiques, l'art, etc.[382]

Un autre avocat de la théorie de la *response-dependency* est Jesse PRINZ qui écrit : « Je défends l'idée que les faits moraux sont *response-dependent* : le mal est simplement ce qui cause de la désapprobation dans une communauté d'êtres moraux. » (PRINZ 2006, p. 29, ma traduction)[383]

La théorie de la *response-dependency* est une forme de théorie de la survenance. Elle semble avoir l'avantage d'apporter plus de précisions sur la nature des propriétés morales ; celles-ci surviendraient sur des états du monde en ce qu'ils entretiennent une certaine relation avec *notre* perception ou évaluation du monde (étant entendu que ces dernières dépendent de notre nature, de la manière dont nous sommes constitués). L'idée de la *response-dependency* impose que les sujets moraux aient conscience de la bonté de certains états de fait pour que ces derniers puissent présenter la propriété du bien.

Quant à l'accès cognitif à cette réalité morale, les auteurs invoquent souvent les émotions ; ces dernières seraient à la fois révélatrices et créatrices de moralité. PRINZ par exemple, pense que les jugements moraux sont des réactions émotionnelles, si bien que la manière dont nous percevons et évaluons les états de fait est déterminée par nos

---

[381] « The integrationist contends that x can only be good for something; x cannot just be good. That does not mean that there are no intrinsic values; it just means that if something is intrinsically valuable, it is so because it has a special connection with a certain sort of agent. » (ROTTSCHAEFER 1998/1997, pp. 219-220)

[382] « In addition to the primarily biologically based value of human fitness there are other, primarily culturally based human values, for instance, science, mathematics and art. » (ROTTSCHAEFER & MARTINSEN 1990, p. 168; pour un développement semblable voir aussi ROTTSCHAEFER 1998/1997, pp. 220-221)

[383] « I argue that moral facts are response-dependent: the bad just is that which causes disapprobation in a community of moralizers. » (PRINZ 2006, p. 29)

dispositions innées à avoir certaines réactions émotionnelles (elles forment notre « plan sentimental »).[384]

La conception que se fait ROTTSCHAEFER de l'accès cognitif à la réalité morale est plus complexe. Selon lui, la manière dont nous percevons et évaluons les états de choses dépend de deux mécanismes : premièrement la manière dont nous sommes constitués, c'est-à-dire nos capacités innées (en particulier les émotions empathiques) et deuxièmement l'apprentissage culturel (ROTTSCHAEFER & MARTINSEN 1990, p. 163).[385] Ces deux facteurs sont non seulement la cause de nos jugements moraux mais ils les justifient du même coup. Ainsi, de même que, dans des conditions idéales nos yeux nous fournissent un accès cognitif fiable au monde physique qui nous entoure, de même certains mécanismes nous permettent de saisir le bien et le mal des objets et états de choses que nous observons (MARTINSEN & ROTTSCHAEFER 1990, p. 159).

A première vue, la théorie de la *response-dependency* permet de mieux définir à la fois la réalité ontologique des propriétés morales et la faculté qui permet de les saisir. Toutefois, je soutiens que ces raffinements conceptuels n'améliorent pas le sort du réalisme moral. Au contraire il semblerait qu'ils ajoutent une dimension d'étrangeté supplémentaire aux propriétés morales. La théorie de la *response-dependency* est tiraillée entre deux extrêmes qu'elle rejette. Elle ne peut pas signifier que les propriétés morales dépendent des attitudes et croyances *individuelles* des gens, car ces dernières ne rempliraient pas les conditions même du réalisme moral et l'on se retrouverait à défendre une forme d'anti-réalisme ; il est donc indispensable que les propriétés *response-dependent* soient imposées de l'extérieur, qu'elles soient indépendantes de la

[384] « Basic moral values may consist in having sentiments associatively linked in long-term memory to specific kinds of actions, abstractly construed. (…) A judgment that some action is wrong counts as erroneous if that action is not an instance of a type towards which we have a sentimental policy. » (PRINZ 2006, p. 35)

[385] Ce second facteur est assez déroutant. En effet, je ne vois pas comment l'apprentissage socioculturel pourrait produire une base suffisamment stable pour permettre de parler de propriétés morales qui possèdent une réalité propre. Sans doute ROTTSCHAEFER a-t-il à l'esprit l'idée que l'apprentissage est un moyen de développer des capacités dont le bon fonctionnement nécessite un certain entraînement (un peu à la manière aristotélicienne d'actualisation de puissances). Le fait qu'il parle d'un sens moral qui doit être éduqué, semble parler en faveur de cette hypothèse (ROTTSCHAEFER & MARTINSEN 1990, p. 160).

pure subjectivité des individus.[386] En bref, si l'on veut assurer une réalité ontologique propre aux propriétés morales postulées par cette théorie, il faut s'assurer que la dimension « réponse » de la propriété soit stable. Or si cette stabilité se résume à des capacités, désirs ou émotions universels aux êtres humains, alors les propriétés morales se trouvent réduites à des propriétés non morales à la manière décrite par des auteurs comme CASEBEER ou ARNHART. Mais ce n'est certainement l'objectif d'un défenseur de la *response-dependency*.[387] Ainsi, sa position vogue entre deux extrêmes sans que l'on comprenne vraiment en quoi elle consiste. Les propriétés postulées ont à la fois une réalité indépendante et dépendante de la contingence humaine. Ici, on ne peut s'empêcher de penser à nouveau à l'argument de l'étrangeté de John MACKIE (1977, p. 38-41 ; voir cet ouvrage, p. 274). Enfin, compte tenu du caractère indéfinissable de ces propriétés morales, la capacité de les découvrir s'imprègne également de mystère et, par là même, devient sujette aux effets corrosifs de l'acide universel (voir p. 280 et suiv.).

6.1.6. Quelques arguments supplémentaires contre le réalisme cognitiviste

Au terme de cette analyse des différentes versions possibles d'une position réaliste cognitiviste, il apparaît qu'aucune n'est acceptable pour un éthicien évolutionniste. Afin de conforter cette analyse, cette section présente deux arguments supplémentaires et des données empiriques incompatibles avec le réalisme. Ces éléments étant de portée générale, ils peuvent également être utilisés en métaéthique traditionnelle.

Puisque l'existence des propriétés morales ne peut être que postulée (à ce propos, voir aussi PUTNAM 2004) ou au mieux rendue plausible, le réaliste cognitiviste se doit de produire la meilleure explication de notre activité morale en termes de faits moraux conçus comme sources causales de nos croyances et de nos actions. Avec BLACKBURN (1993, p. 180) et bien d'autres, je suis sceptique quant au pouvoir explicatif du postulat de l'existence des propriétés morales. Les sections précédentes ont montré les limites de ce postulat. Les conclusions auxquelles je suis parvenue peuvent être renforcées par

[386] ROTTSCHAEFER admet cela puisqu'il affirme que la vérité d'une croyance morale est indépendante de la souscription individuelle à cette croyance (1998/1997, p. 165).

[387] ROTTSCHAEFER par exemple refuserait d'emprunter le chemin réductionniste ; il exclut explicitement la possibilité d'identifier les phénomènes moraux à des phénomènes non moraux (1998/1997, p. 165).

deux objections souvent présentées contre le réalisme moral. Si l'extension des *concepts* moraux était déterminée par les *faits* moraux, nous serions confrontés à au moins deux situations embarrassantes. Premièrement, nous serions forcés d'affirmer que deux personnes ne peuvent pas avoir une pleine compréhension d'un état de fait sans s'accorder sur la valeur morale de cet état de fait. En d'autres termes, un réaliste cognitiviste naturaliste doit rejeter la possibilité que quelqu'un soit en désaccord avec une autre personne au sujet de l'application correcte d'un terme moral sans qu'il y ait une mécompréhension entre eux au sujet du fait en question. Or notre intuition nous laisse précisément croire que cela est possible (HORGAN & TIMMONS 1992). Deuxièmement, de manière tout aussi embarrassante, le réalisme cognitiviste nous impose l'impossibilité de penser et parler d'une valeur morale (par exemple la justice) dans un monde qui serait totalement dénué de la propriété correspondante. Cela paraît également aberrant (HORGAN & TIMMONS 1992 ; BLACKBURN 2000/1998, p. 203).

Pour terminer, il vaut la peine de noter que le projet de fonder la portée universelle des principes moraux (vers lequel tend toute théorie métaéthique réaliste) est peu compatible avec les données empiriques récoltées en psychologie et anthropologie. Ces dernières infirment l'idée que les êtres humains présentent des réponses évaluatives stables face à des situations moralement pertinentes.

Il semblerait que nos jugements moraux soient très largement dépendants des contextes historiques et sociaux dans lesquels ils sont produits. Par exemple, l'homosexualité est largement acceptée dans certaines régions du monde alors qu'elle est proscrite dans d'autres régions. Dans le même ordre d'idée, les travaux de Richard NISBETT et Dov COHEN (1996) indiquent que dans les cultures de l'honneur (notamment dans les Etats-Unis du Sud), les gens acceptent aisément les actes de violence pour peu qu'ils soient accomplis dans le but de sauver son honneur; dans certaines circonstances, ils les considèrent même comme recommandables. Cela ne correspond évidemment pas à la manière dont les citoyens des Etats-Unis du Nord ou d'Europe du Nord jugent les actes de violence motivés par la défense de l'honneur.

En plus de la relativité des normes défendues dans diverses sociétés et époques, il semblerait que beaucoup d'êtres humains soient intuitivement relativistes. Récemment, Daniel KELLY et ses collègues ont questionné des sujets sur la pratique de la flagellation des mauvais marins en mer. La moitié d'entre eux trouvent que cette pratique était acceptable il y a 300 ans et seulement 10 % pensent qu'elle est admissible de nos jours (KELLY *et al.* 2007). Dans le même ordre d'idées, les travaux de Karen HUSSARD sous

la direction de Paul HARRIS (données non publiées) sur les jugements moraux des enfants végétariens révèlent qu'une grande partie d'entre eux justifient leurs habitudes de nourriture en invoquant qu'il est mal de causer du tort aux animaux. En revanche, ils ne condamnent pas les personnes qui mangent de la viande. Si ces enfants saisissaient vraiment le mal dans l'action de tuer un animal, ils ne pourraient pas accepter l'attitude des non-végétariens (il en va de même des jugements sur la flagellation).

Ces données parlent en faveur de l'idée que l'extension des *concepts* moraux n'est pas déterminée par des *faits* moraux, que ce soit par perception directe des propriétés morales ou par influence causale des faits moraux sur la conception que l'on s'en fait.

6.2. L'antiréalisme

Le rejet du réalisme cognitiviste conforte l'analyse descriptive de l'activité morale présentée au chapitre précédent. Cette analyse s'accorde davantage avec la thèse antiréaliste projectiviste selon laquelle nos concepts de bien ou de mal moral ne correspondent pas à une réalité extérieure à nos croyances et attitudes; nous projetons les valeurs sur le monde.

La thèse antiréaliste s'accompagne forcément du non-cognitivisme ; puisqu'il n'y a pas de réalité morale, nous ne possédons pas de moyens cognitifs pour y accéder.[388] Au niveau sémantique, l'antiréalisme implique que les énoncés moraux ne sont ni vrais ni faux et tirent leur sens uniquement des conditions dans lesquelles ils sont assertés.[389]

[388] On dit parfois que les défenseurs de la théorie de l'erreur (MACKIE, RUSE etc.) sont cognitivistes antiréalistes car selon eux, les être humains sont convaincus de l'existence d'une réalité morale et de leur capacité d'y accéder (mais ils se trompent si bien que toutes leurs assertions morales sont fausses). Toutefois, cette affirmation est contradictoire à moins de comprendre la notion de cognitivisme au sens psychologique. Il est clair que des auteurs comme MACKIE nient le cognitivisme en tant que description de ce dont les êtres humains sont réellement capables ; et lorsque les théoriciens de l'erreur affirment que toutes les assertions morales sont fausses, c'est en vertu du fait qu'elles sont formulées sous forme de propositions par des *êtres humains* convaincus (à tort) qu'elles peuvent être vraies ou fausses.

[389] Si l'on veut défendre la possibilité d'une vérité morale dans le cadre d'une théorie antiréaliste, alors il faut assouplir la notion même de vérité et ne plus la considérer comme une correspondance entre une assertion et la réalité.

L'antiréalisme projectiviste se décline sous différentes formes, les plus connues étant l'émotivisme, l'expressivisme et la théorie de l'erreur. Elles se distinguent entre elles par la manière dont elles conçoivent i) le statut sémantique des assertions morales et ii) la nature des états mentaux corrélés à ces assertions.

La première problématique traite de la nature des assertions morales. Sont-elles propositionnelles (théorie de l'erreur) ou s'agit-il de simples expressions d'attitudes ou d'émotions (expressivisme, émotivisme) ? A défaut de vérité au sens de correspondance entre l'assertion et la réalité, peut-on dire que ces assertions sont correctes ou incorrectes, justifiées ou injustifiées ?

La seconde problématique concerne la question de savoir si les états mentaux liés aux assertions morales sont de simples attitudes ou émotions (émotivisme), un engagement non propositionnel en faveur de certaines émotions (expressivisme), ou un type particulier de croyances (théorie de l'erreur).

Compte tenu de la division des champs de l'éthique présentée au chapitre 4 (section 4.1), les questions de justification des assertions morales sont traitées dans le domaine de l'éthique normative et celles relatives à la nature des états mentaux corrélés aux assertions morales relèvent davantage du domaine de l'éthique descriptive. Ainsi, le choix d'une position métaéthique antiréaliste dépend grandement des idées développées à ces deux autres niveaux de l'éthique.

Dans ce qui suit, je vais brièvement esquisser et rejeter l'émotivisme, l'expressivisme et la théorie de l'erreur avant d'adopter une position projectiviste similaire à la dernière théorie, mais qui accorde peu d'importance à la thématique de l'erreur.

6.2.1. Le rejet de l'émotivisme

L'émotivisme a été défendu de manière paradigmatique par Alfred AYER (1946/1936) et Charles STEVENSON (1937).[390] Selon ces auteurs, en énonçant un

[390] On trouve déjà les prémisses de ce type de pensée dans l'histoire de la philosophie. Baruch de SPINOZA par exemple, soutenait l'exact contraire de la position platonicienne en écrivant : « Nous ne nous efforçons à rien, ne voulons, n'appétons ni ne désirons aucune chose, parce que nous la jugeons bonne ;

jugement moral, les gens ne feraient qu'exprimer un sentiment ou une attitude d'approbation ou de désapprobation associée à une prescription.

Cette position métaéthique semble toutefois entrer en contradiction avec mon analyse descriptive de l'activité morale (section 5.2). Selon mon tableau affectif, ce n'est pas simplement en exprimant une émotion que l'on adhère à des valeurs morales ou que l'on tient consciemment compte d'autrui pour décider de ce qui est bien et mal. Je ne peux donc pas retenir une telle position, d'autant plus qu'il y a une foule d'autres raisons de la rejeter. L'émotivisme a en effet été l'objet d'attaques cinglantes et répétées depuis sa formulation dans les années trente (voir par exemple B. WILLIAMS 1973; M. SMITH 1986). Voici quelques objections. Tout d'abord, cette position ne permet pas de distinguer entre les réactions émotionnelles ou attitudes qui ont une autorité morale et celles qui n'en ont pas (B. WILLIAMS 1973). Ensuite, si les assertions morales se résument effectivement à de simples expressions d'attitudes ou émotions, l'émotivisme peine à expliquer le fait que des assertions morales puissent faire l'objet de débats rationnels[391] et le fait que de réels désaccords puissent survenir entre deux personnes au sujet de la valeur morale d'une situation (ACTON 1936 ; GEACH 1965).[392] Enfin l'émotivisme n'admet pas la possibilité d'être horrifié par un état de fait (par exemple en apprenant que son voisin a vendu au boucher la viande de son cheval mort par accident) sans pour autant le considérer comme moralement condamnable.

---

mais, au contraire, nous jugeons qu'une chose est bonne parce que nous nous efforçons vers elle, la voulons, appétons et désirons. » (*Ethique* III, prop. 9, scolie)

[391] Par exemple, les énoncés moraux conditionnels (donc non assertifs) semblent poser problème pour l'émotivisme. Est-ce que l'énoncé « Si tuer des enfants innocents est *mal*, payer quelqu'un pour qu'il tue des enfants innocents est mal » peut être compris comme l'expression d'une émotion ? (GEACH 1965). Notons qu'en dépit de sa popularité, cet argument n'est probablement pas le meilleur pour contrer l'émotivisme (à ce propos voir la note suivante).

[392] Il y a peut-être une réponse à cet argument. Simon BLACKBURN fait remarquer qu'un émotiviste peut parfaitement bien exprimer une attitude en réponse à l'attitude de quelqu'un d'autre. De plus, il peut également exprimer la conviction que si une attitude est de mise, alors une autre l'est également, et ainsi de suite. Il est possible d'imaginer un débat de cette sorte entre émotivistes (1993, p. 19).

6.2.2. Le rejet de l'expressivisme

C'est pour pallier aux faiblesses de l'émotivisme tout en maintenant l'idée que les assertions morales ne sont que des expressions d'attitudes, qu'Allan GIBBARD (2002/1990) et Simon BLACKBURN (1993 ; 2000/1998) ont développé une théorie plus sophistiquée appelée « expressivisme ». Selon cette approche, les assertions morales seraient des états mentaux d'un type particulier ; ce seraient des expressions d'acceptation ou approbation d'une réaction émotionnelle face à une situation moralement pertinente.

L'expressivisme ne s'accorde pas plus que l'émotivisme avec mon tableau affectif : aucune des trois formes de jugements de valeur relatés dans ce tableau ne se résume à un état mental (acceptation, approbation) au sujet d'un état mental (réaction émotionnelle). En outre, de même que l'émotivisme, l'expressivisme doit faire face à des objections sérieuses. Justin D'ARMS et Daniel JACOBSON remarquent par exemple que l'on peut très bien juger qu'un sentiment de culpabilité est approprié face à une action, sans pour autant considérer cette action comme mauvaise.[393] Or il s'agit d'une attitude inconcevable dans le cadre d'une théorie expressiviste (1994, p. 742). A l'image de l'émotivisme, l'expressivisme éprouve également des difficultés à différencier les assertions morales des assertions non morales. Pour y échapper, un défenseur de cette position pourrait postuler que, dans les situations morales, notre acceptation ou approbation elle-même est de nature *morale* ; mais cette solution semble introduire une circularité dans la théorie (D'ARMS & JACOBSON 2000). Une autre solution est d'établir une distinction entre les émotions de nature morale et non morale ;[394] mais nous avons vu à la section 5.3 qu'une telle distinction ne semble pas viable. Enfin, recourir à une explication métacognitive à la manière expressiviste semble à la fois intuitivement peu crédible (il est difficile d'admettre que les assertions morales fassent preuve d'une telle sophistication) et peu utile, car il est possible de défendre une théorie antiréaliste qui

[393] Leur exemple est celui d'une personne qui se résout à placer sa mère, atteinte d'une maladie débilitante, dans une institution spécialisée.

[394] Allan GIBBARD s'est lancé dans cette entreprise en choisissant la colère et la culpabilité comme émotions morales par excellence.

considère les assertions morales simplement comme des propositions (et non des expressions d'états d'esprits ou attitudes).

6.2.3. Les limites de la théorie de l'erreur

Parmi les différents classiques antiréalistes, la théorie de l'erreur est la position la plus compatible avec mon tableau affectif. L'auteur de référence de cette théorie est sans aucun doute John MACKIE (1977), mais c'est dans le cadre de l'éthique évolutionniste que la théorie de l'erreur a été pleinement développée (RUSE 1998/1986 ; SOMMERS & ROSENBERG 2003 ; JOYCE 2000). Les défenseurs de cette position rejettent les versions antiréalistes qui conçoivent les assertions morales comme des expressions d'états d'esprit (émotivisme, expressivisme). Ils préfèrent s'en tenir à une interprétation plus simple en termes de croyances à contenu propositionnel. D'autre part, ils sont convaincus que si l'on rejette ces versions antiréalistes en plus du réalisme moral, il reste une seule voie possible : la théorie de l'erreur. En cela je pense qu'ils ont tort ; mais avant de le montrer, il me faut présenter leur point de vue.

En deux mots, selon la théorie de l'erreur, les gens croient en l'objectivité de leurs jugements moraux mais ils se trompent car les propriétés morales n'existent pas. La notion d'objectivité est ici comprise au sens où les gens sont réalistes, c'est-à-dire croient en l'existence d'une réalité morale qui existe de manière indépendante de leurs croyances et attitudes. Ainsi, du point de vue sémantique, toutes nos assertions morales sont fausses.

Selon les versions évolutionnaires de la théorie de l'erreur, cette croyance fausse (ou illusion) serait inscrite dans nos gènes.[395] Plus précisément, elle aurait été sélectionnée parce qu'elle a pour effet de renforcer les comportements coopératifs adaptatifs (RUSE 1998/1986, pp. 253-256). En d'autres termes, les assertions morales et notre croyance en leur validité objective sont très utiles puisqu'elles nous incitent à adopter des comportements coopératifs ; en revanche, elles sont toutes fausses. De cela,

[395] « In particular, the evolutionist argues that, thanks to our science, we see that claims like 'You ought to maximize personal liberty' are no more than subjective expressions, impressed upon our thinking because of their adaptive value. In other words, we see that morality has no philosophically objective foundation. It is just an illusion, fobbed off on us to promote biological 'altruism'. » (RUSE 1998/1986, p. 102)

les théoriciens de l'erreur concluent que la morale ne peut pas être fondée, c'est-à-dire qu'aucun énoncé moral ne peut être justifié.

Pour soutenir leur point de vue, les théoriciens de l'erreur évoquent souvent un argument d'économie. Voici la version de Tamler SOMMERS et Alex ROSENBERG :

> « La meilleure explication – variation aveugle et sélection naturelle – de l'émergence de nos croyances éthiques ne requiert pas que ces croyances puissent être vraies ou fausses (…) ; ainsi, le nihiliste [autre terme qu'ils utilisent pour signifier : théoricien de l'erreur] peut se contenter d'ajouter le principe scientifique non controversé selon lequel, si la meilleure théorie pour expliquer pourquoi les gens croient P ne requiert pas que P soit vrai, alors il n'y a aucune raison de penser que P est vrai » (SOMMERS & ROSENBERG 2003, p. 667, ma traduction)[396]

En d'autres termes, si la *croyance* en une réalité morale remplit la fonction de rendre les gens coopératifs et altruistes, alors il n'est pas nécessaire que l'évolution ait créé une réalité morale indépendante des croyances des gens.

A mes yeux, la théorie de l'erreur comporte une face séduisante (l'antiréalisme ; la compréhension des assertions morales en termes de croyance) et une autre moins appréciable (la notion d'illusion ; l'impossibilité de justifier les assertions morales). Dans ce qui suit, j'aimerais montrer qu'il est possible de maintenir la première face et de se départir de la seconde.

La théorie de l'erreur postule que les êtres humains croient en une réalité morale extérieure. Mais ce postulat peut être remis en question sur la base de simples données empiriques. Plus haut (p. 298) nous avons vu que les gens peuvent être relativistes dans leur jugements (expériences sur la flagellation et le végétarisme). D'autre part, sur la base de tests empiriques destinées à sonder les intuitions réalistes des gens, Shaun NICHOLS (2004, p. 177) parvient à la conclusion que nous ne sommes pas forcément réalistes, même si nous présentons une tendance en ce sens (voir aussi STICH & WEINBERG 2001, RYAN 1997). Je ne cherche pas à nier que les gens croient en la portée

[396] « The best explanation – blind variation and natural selection – for the emergence of our ethical belief does not require that these beliefs have truth-makers (…); the nihilist need only add the uncontroversial scientific principle that if our best theory of why people believe P does not require that P is true, then there are no grounds to believe P is true. » (SOMMERS & ROSENBERG 2003, p. 667)

intersubjective de leurs assertions morales ; mais il me semble que cette portée est plus ou moins large en fonction de la force de leurs réactions émotionnelles. Au fond, il ne s'agit de rien de plus que l'objectivité psychologique telle qu'elle a été définie à la section 5.2.1.vi. Vouloir y ajouter la croyance en une réalité morale extérieure semble inutile voire même redondant.[397] Ironiquement, si l'on pousse l'argument d'économie jusqu'au bout, on constate que l'évolution n'a pas non plus besoin de créer en nous une illusion pour nous rendre coopératifs. Au contraire, il n'y a aucune raison de croire que les êtres humains aient évolué de manière à faire l'expérience de préférences et d'obligations imposées de l'extérieur ; ce serait attendre trop de l'évolution (GEIGER 1992).

D'autre part, l'affirmation selon laquelle la morale ne peut pas être fondée perd également de son attrait lorsque l'on prend conscience de la conception rigide de la justification morale sur laquelle elle repose. La théorie de l'erreur postule que pour qu'une assertion soit justifiée, elle doit être démontrée vraie ; de plus, cette vérité est comprise comme une correspondance entre une croyance et une réalité externe aux croyances et attitudes du sujet. Dans le chapitre suivant (section 7.4), je tâcherai de montrer qu'il existe d'autres moyens de peser les jugements moraux et d'accéder à un certain degré de justification.

En résumé, il est possible de maintenir le projectivisme tout en faisant l'économie de l'illusion et de l'impossibilité de justifier nos assertions morales.

6.2.4. Défense d'un projectivisme simple

Après élimination du réalisme moral et des trois courants antiréalistes majeurs, la voie métaéthique qui me paraît la plus prometteuse est une forme de théorie de l'erreur désamorcée, c'est-à-dire exempte de la thématique de l'erreur et de l'impossibilité de fonder la morale.

---

[397] Il me semble que David LAHTI touche un point sensible lorsqu'il écrit: « A major problem with the error theorist's hypothesis, however, is evident before any such analysis is performed : the hypothesis simply makes morality redundant. The evolutionary ethics error theorist enlists the moral law to take up a post that is already occupied. The caring sentiments natural to our species (intentional altruism) can be felt and discussed without reference to apparently objective moral guidelines. » (LAHTI 2003, p. 643-644). Pour une argumentation similaire, voir COLLIER & STINGL (1993, pp. 53-54). Pour une réponse à cette critique, voir JOYCE (2006, p. 115).

La position que je défends est un antiréalisme projectiviste selon lequel les assertions morales sont des croyances d'un certain type : des projections de valeurs sur le monde assorties d'une impression de validité intersubjective. Cette dernière est causée par la puissance de nos réactions émotionnelles en conjonction avec l'activité de l'empathie égocentrique (5.2.1.vi p. 213). Ainsi, les assertions morales sont objectives dans la mesure où les gens les considèrent comme telles. De plus, la moralité comporte un autre aspect objectif : l'objectivité de fait (section 5.2.1.vi). Notre nature humaine, plus précisément les biais psychologiques qui ont évolué au cours de notre histoire phylogénétique, endiguent le spectre de nos réactions émotionnelles et dirigent une partie de nos comportements en direction de l'entraide, coopération et altruisme. Conjointement, ces facteurs œuvrent à ce que la pensée morale s'impose à l'esprit des gens, influence leur comportement, et en fin de compte, cimentent les relations à l'intérieur et entre les communautés humaines.

Il se pourrait que les options de l'antiréalisme et de l'objectivité limitée prises dans ce chapitre occasionnent une crainte du relativisme moral. En réalité, le relativisme moral n'est pas aussi dangereux qu'il n'y paraît, sous réserve que l'on ne le conçoive pas dans un sens radical et qu'il ne mette pas en péril les deux conditions suivantes : les gens disposent de moyens efficaces pour accorder leurs assertions morales et ces dernières peuvent être justifiées. J'ai déjà répondu par la positive à la première condition (section 5.2.1.vi). Quant à la seconde, elle fera l'objet du chapitre suivant. Je tâcherai d'y montrer que le meilleur angle d'approche de la moralité est d'ordre psycho-épistémique : ce qui importe, c'est ce que les gens pensent et comment ils agissent ; or il se trouve que malgré leurs tendances relativistes, ils pensent que les normes et valeurs auxquelles ils adhèrent peuvent être fondées et ont une valeur intersubjective. Dans ce contexte, il est possible de parler de justification de nos assertions morales.

Conclusion

Au terme de ce chapitre, j'espère avoir pu montrer que l'adoption d'une perspective évolutionnaire exerce un impact non négligeable sur le genre de positions métaéthiques que l'on peut défendre. Elle nous impose de rejeter toute théorie réaliste cognitiviste ; elle nous fait comprendre que notre univers est évaluativement neutre, même s'il contient des créatures qui, elles, sont capables de produire des jugements de

valeur. Ces derniers ne correspondent pas à quelque essence morale indépendante de nos attitudes et croyances.

En outre, faisant appel à mon analyse descriptive de la moralité, j'ai soutenu que les jugements de valeur (et les assertions morales en général) sont l'expression de la manière dont nous nous représentons le monde. Ainsi nous projetons les valeurs sur le monde. Mais cela ne se fait pas au hasard : la moralité est construite sur la base de capacités et biais psychologiques précis qui guident notre pensée et notre activité morales.

La dernière tâche à remplir pour achever mon projet de défense d'une éthique évolutionniste est de montrer que le rejet d'un fondement ontologique des normes, valeurs et jugements moraux ne compromet pas la possibilité de la justification morale. Cette question relève de l'éthique normative à laquelle est consacré le dernier chapitre de cet ouvrage.

7. Point de vue évolutionnaire sur l'éthique normative

Les réflexions menées en éthique descriptive et en métaéthique concluent à l'impossibilité d'un fondement ontologique des valeurs et des normes. Cela nous pousse à admettre qu'en matière de morale, la recherche de la vérité est vaine. En revanche, la question de la justification morale garde toute son importance. Une des tâches les plus importantes du domaine de l'éthique normative consiste à élaborer et justifier les *éléments de base* de la morale, c'est-à-dire les valeurs, les normes de premier ordre et autres notions fondamentales sur lesquelles repose l'ensemble d'une théorie morale. Ce chapitre porte sur la manière dont les théories évolutionnistes et les données scientifiques peuvent aider à justifier ces éléments de base.

Beaucoup de tenants de l'éthique évolutionniste pensent que les valeurs et normes morales peuvent trouver un ancrage dans les données factuelles ; elles dépendraient directement de la manière dont la nature nous a façonnés (RICHARDS 1986 ; ROTTSCHAEFER 1998 ; ARNHART 1998 ; COLLIER & STINGL 1993).

D'autres penseurs issus de divers disciplines s'opposent à l'éthique évolutionniste (NAGEL 1983/1978 ; O'HEAR 1997 ; HUXLEY 1893 ; ALEXANDER 1987, 1979 ; G. WILLIAMS 1988 ; 1993 ;[398] GOULD 1999 ; DAWKINS 1996/1976). Pour eux, les sciences comme la biologie, la neurologie ou la psychologie, n'ont pas de contribution fondationelle ou justificatrice au niveau normatif ; l'éthique doit être séparée de la nature. Ainsi Richard DAWKINS écrit :

> « Si vous voulez, comme moi, construire une société dans laquelle les individus coopèrent généreusement et sans égoïsme pour réaliser le bien commun, vous ne pouvez attendre beaucoup d'aide de la Nature. Essayons de comprendre ce vers quoi tendent nos gènes, c'est-à-dire l'égoïsme, parce qu'il se pourrait alors que nous ayons au moins une chance de déjouer leurs plans et d'atteindre ce à quoi aucune autre espèce n'est jamais parvenue, devenir un individu altruiste. » (DAWKINS 1996/1976, p. 19)

Dans le même ordre d'idée, le philosophe néo-kantien Thomas NAGEL (1983/1978) considère l'éthique comme une recherche théorique sur la pratique et les motifs des

[398] Des auteurs comme Richard ALEXANDER ou George WILLIAMS vont même jusqu'à affirmer qu'il est préférable de ne pas suivre nos instincts biologiques.

gens que l'on peut aborder par des méthodes rationnelles et qui possède des critères internes de justification et de critique. Du fait que son domaine est très différent de celui des autres sciences (elle concerne les raisons et la motivation à agir) et qu'elle « possède des critères internes de justification et de critique » (p. 167), l'éthique ne peut pas être traitée à l'aide de méthodes scientifiques.[399] A cela, il ajoute :

> « [L'éthique] est le résultat d'une capacité humaine à soumettre des schèmes de motivation et de comportement pré-réflexifs innés ou conditionnés à critique et à révision, et à créer de nouvelles formes de conduite. La capacité à faire cela a sans doute quelque fondement biologique, même s'il ne s'agit que d'un effet secondaire d'autres développements. Mais l'histoire de l'exercice de cette capacité et l'application continuelle qu'elle ne cesse de faire de ses propres produits à la critique et à la révision ne font pas partie de la biologie. » (NAGEL 1983/1978, p. 171)

En d'autres termes, dès lors que la raison est apparue, les hommes ont été capables de se dégager de leurs déterminants génétiques et d'adopter des comportements moraux. L'éthique serait donc uniquement un effet de la raison. La seule concession à laquelle est prêt NAGEL concerne l'existence de quelques obstacles psychologiques et sociaux dont certains peuvent avoir des causes biologiques ; mais il les considère comme tout à fait surmontables (1983/1978, p. 172). Au fond, il est persuadé que notre nature biologique ne nous donne strictement aucune indication sur les attitudes morales à adopter.

Je tâcherai de montrer qu'aucun de ces deux partis opposés n'est viable et qu'il faut préférer une position médiane. Pour commencer, je mettrai en évidence le fait que toutes les théories morales sont confrontées à la difficulté de justifier les éléments de base sur lesquels elles reposent et qu'aucune des solutions issues de la tradition morale non évolutionniste n'est pleinement convaincante. Je présenterai et rejetterai ensuite quelques projets de justification issus de l'éthique évolutionniste ; il apparaîtra que l'on ne peut pas recourir sans autre aux données scientifiques et aux théories évolutionnistes pour fonder un système moral. Je poursuivrai en arguant que ces mises en garde n'imposent pas d'adopter une position défaitiste qui abandonne toute tentative de fonder

[399] NAGEL considère que la moralité est uniquement affaire de raison et qu'elle est susceptible de progrès au même sens que la physique ou les mathématiques.

les normes morales. Le chapitre se terminera sur l'exposé d'une forme d'éthique évolutionniste qui consiste à recourir au « sens commun renforcé ».

7.1. Le problème du fondement des éléments de base des théories morales

Tout système de philosophie morale repose sur un certain nombre de présupposés difficilement justifiables. Il s'agit essentiellement de valeurs conscientes et de normes de premier ordre ; ces dernières peuvent être interprétées en termes de valeurs. Par exemple, il est possible d'interpréter le principe utilitariste (BENTHAM 1948/1789 ; MILL 1998/1861) comme reposant sur une valeur morale par excellence : la réalisation de la plus grande quantité[400] de plaisir ou de bonheur en comparaison des souffrances et peines (il s'agit en l'occurrence d'une valeur relationnelle). Pour KANT (1997/1785 ; 1997/1788), la valeur par excellence est le fait d'être de bonne volonté, c'est-à-dire de se contraindre à prendre uniquement des décisions dont on peut vouloir en même temps qu'elles valent pour toute personne rationnelle. [401]

Nous avons vu au chapitre 5 (section 5.2.1.v) que nous choisissons généralement nos normes et valeurs en fonction de nos sentiments et des influences de notre environnement. Mais cela ne nous dit rien sur la question de leur *justification*. Selon la chaîne de justification décrite à la section 5.4.3, les normes de second ordre reposent sur des normes morales de premier ordre et ultimement sur des valeurs. Ces éléments de base étant difficilement justifiables, ils sont souvent acceptés a priori ; on les tient pour si fondamentaux qu'ils ne requièrent aucune justification en amont. Mais beaucoup de philosophes ne se contentent pas de ce procédé et tentent de les justifier ou au moins d'établir une hiérarchie[402] entre eux afin de pouvoir gérer les situations où différentes

[400] Selon les modèles, il s'agit d'une agrégation quantitative (BENTHAM) ou qualitative (MILL).

[401] Une position hédoniste radicale définira le bonheur personnel comme valeur par excellence ; mais nous avons vu que selon les critères de moralité définis à la section 5.4.1, cette forme d'hédonisme ne peut pas être considérée comme morale (voir section 5.6 p. 265). Il s'agirait en fait plutôt d'une doctrine d'intérêt rationnel.

[402] En effet, la crédibilité d'un système moral tient en bonne partie à la commensurabilité des normes et valeurs fondamentales sur lesquelles il repose. A cet égard, soit on accepte une seule valeur morale et sa norme de premier ordre (c'est par exemple la voie choisie par les utilitaristes), ou on élabore un système

normes ou valeurs fondamentales entrent en conflit direct ou prescrivent des actions incompatibles entre elles. A cet effet, différentes voies peuvent être empruntées. Ci-dessous, je vais tenter d'en établir une liste plus ou moins exhaustive, même si les limites entre ces différentes options sont assez arbitraires ; les systèmes philosophiques intègrent souvent plusieurs approches.

Pour fonder les normes et les valeurs, on peut simplement faire référence au sens commun, c'est-à-dire à l'opinion des gens ordinaires. Mais du point de vue philosophique, ce procédé est peu probant. D'une part, la liste des opinions du sens commun ne peut pas être facilement établie ; les gens véhiculent souvent des croyances qu'ils ont des difficultés à concevoir de manière explicite dans leur esprit. D'autre part, il est aisé de trouver des contre-exemples qui portent à croire que le sens commun défend en réalité la thèse opposée. Par exemple, s'il s'agit de statuer sur la peine de mort pour les tueurs en série, le sens commun peut être invoqué à la fois en faveur et contre cette pratique. Notons que je ne nie pas ici l'intérêt de faire appel au sens commun ; au contraire, il me semble que cette attitude est essentielle en éthique. Par contre il faut admettre qu'il n'est pas suffisant pour fonder un élément de base de la morale.

Une autre solution serait de recourir à l'intuition (MOORE 1903), ou à des émotions (SCHELER 1955/1913 ; TAPPOLET 2000 ; KRISTJANSSON 2002 ; GOLDIE 2007 ; DÖRING, 2007) censées nous fournir un accès épistémique aux propriétés morales. Mais pour être crédible, tout défenseur d'une telle position doit admettre que l'intuition ou les émotions sont de plus ou moins bons révélateurs de propriétés morales ; ces dernières jouent le rôle de critère d'adéquation de l'intuition ou des émotions. Par exemple, l'intuition qui nous fait considérer que notre amie anorexique devrait valoriser sa santé plutôt que sa beauté n'est adéquate que si la santé a effectivement plus de valeur que la beauté. De même, une émotion d'indignation morale sera uniquement appropriée si la situation qui nous indigne est effectivement immorale. Ainsi cette approche repose sur une position réaliste en métaéthique, laquelle a été rejetée au chapitre 6 ; je ne pense pas

---

de hiérarchisation (qui opère soit de manière absolue, soit en fonction du contexte) afin d'éviter les situations conflictuelles.

qu'il soit possible de montrer que des propriétés normatives appartiennent de manière intrinsèque à des états de fait.[403]

Une autre voie qui n'implique pas le réalisme métaéthique consiste à s'appuyer sur une conception kantienne de la raison pratique, dont par exemple, la bonne volonté est l'expression. Toutefois, adopter une position rationaliste, c'est-à-dire faire reposer tous les énoncés moraux sur la raison pratique conçue comme détachée de toute influence mondaine (KANT 1997/1788; 1997/1785 ; NAGEL 1970) paraît de plus en plus tenir d'un vieux rêve anthropocentrique des siècles passés.[404] En effet, indépendamment du fait qu'il n'est pas évident de comprendre en quoi consiste exactement cette raison pratique, nous avons vu au chapitre 5 (section 5.2) qu'un nombre grandissant de données nous font découvrir à quel point le choix de nos valeurs et normes dépend de nos réactions émotionnelles. Dès lors, l'idée même d'une activité morale entièrement détachée des émotions et de toute contingence naturelle qui influence ces émotions paraît hautement sujette à caution.[405]

Une autre manière de fonder les éléments de base de la moralité pourrait être de recourir à un acte de foi religieuse. Par exemple, le fait que Dieu ait transmis les dix commandements à Moïse serait une garantie de leur validité. Toutefois, indépendamment du fait que tout penseur non religieux reste dubitatif face à cette manière de procéder, nous avons vu à la section 5.4.3 que le recours à l'autorité

---

[403] De plus, devant la difficulté de décider quelles intuitions ou émotions sont adéquates et lesquelles ne le sont pas, la plupart de ces théories se tournent vers d'autres voies de justification de leurs normes et valeurs. Selon Max SCHELER (1955/1913) par exemple, on accède aux valeurs par un type particulier d'expérience émotionnelle qui ressemble à une forme de connaissance et qui nous permet de hiérarchiser les valeurs ; mais au-delà de cette affirmation, il cherche à montrer que cette hiérarchie se fonde dans la nature humaine elle-même.

[404] Mentionnons également le traditionnel problème de la motivation à l'action que les défenseurs d'une position rationaliste en morale peinent à résoudre. Comme le remarquait déjà HUME (1946/1740), on voit mal comment de pures raisons, par définition détachées des passions, peuvent exercer un pouvoir motivant à l'action morale. Dans le même ordre d'idée, j'ai argumenté à la section 5.2.2 que la motivation est uniquement déclenchée par les sentiments.

[405] Philippa FOOT argumente dans le même sens: « Kant was perfectly right in saying that moral goodness was goodness of the will ; the idea of practical rationality is throughout a concept of this kind. He seems to have gone wrong, however, in thinking that an abstract idea of practical reason applicable to rational beings as such could take us all the way to anything like our own moral code. For the evaluation of human action depends also on essential features of specifically human life. » (2001, p. 14)

religieuse implique que l'on sort du domaine moral ; les normes et valeurs fondées de cette manière sont des normes d'autorité et non des normes morales.

Une autre voie est de faire appel à des critères de justification. Les plus connus sont les critères épistémiques comme la cohérence (une proposition est justifiée si elle est cohérente avec les autres propositions du système ; voir GOLDMAN 1990 ; RYAN 1997), l'équilibre réfléchi (ajustement mutuel de nos jugements en un tout cohérent en prenant en considération tous les éléments disponibles ; voir RAWLS 1997/1971 ; DANIELS 1996) ou l'impartialité (A. SMITH 2003/1759). Mais on peut également fonder les normes et valeurs sur des critères pragmatiques. John DEWEY (1994) est un défenseur remarquable de ce genre de position ; pour lui, une norme est justifiée si elle est effective. Nous verrons plus loin que ces critères sont extrêmement utiles à condition de ne pas leur accorder le monopole de la justification. Je me contente pour l'instant de signaler deux de leurs limites. Premièrement, ils sont appliqués dans des contextes particuliers et ne justifient qu'en fonction de ces contextes ; ainsi l'étendue de leur pouvoir justificatif est limitée. Deuxièmement ils ne font que déplacer le problème de la justification car ils nécessitent eux-mêmes une justification.

En définitive, toutes les voies proposées jusqu'ici comportent leurs limites. J'en viens maintenant à la dernière méthode de justification des éléments de base d'un système moral, celle qui est paradigmatique de l'éthique évolutionniste : il s'agit de recourir à des données factuelles. Cette approche peut emprunter au moins trois voies compatibles entre elles. La première consiste à faire appel à une certaine conception de la nature humaine, la seconde à s'appuyer sur une analyse de la fonction de la moralité elle-même et la troisième repose sur un examen de ce qui est adapté. Voici quelques précisions sur ces trois options.

En s'inspirant généralement d'ARISTOTE qui a développé son *Ethique à Nicomaque* autour de la notion d'essence humaine, un bon nombre de philosophes font reposer leur système moral sur une certaine conception de la nature humaine. Parmi ceux-ci, certains trouvent leur inspiration dans les écrits d'autres philosophes (par exemple FOOT 2001 ; 2002 ; SCHELER 1955/1913). Mais c'est une manière un peu timide de procéder. D'autres auteurs prennent connaissance et s'inspirent des données empiriques et de théories évolutionnistes sur le fonctionnement de l'être humain ; les systèmes qui en découlent peuvent être classés dans le courant de l'éthique évolutionniste (par exemple ROTTSCHAEFER 1998/1997 ; RICHARDS 1986 ; ARNHART

1998 ; CASEBEER 2003). Dans les deux cas, la stratégie adoptée revient à définir la nature humaine et montrer que tous les êtres humains partagent certains buts, désirs ou besoins essentiels, puis à déduire les normes et valeurs morales à partir des conditions de réalisation de ces buts, désirs ou besoins. On accorde ainsi à la nature humaine un pouvoir justifiant.

On peut également faire appel à la fonction de la morale. Nous avons vu au chapitre 5 (sections 5.1 et 5.5) que la morale n'est pas un résultat direct de la sélection naturelle. On ne peut donc pas lui attribuer de fonction biologique au sens étiologique du terme (voir p. 25). Par contre il peut s'avérer que même si elle n'a pas été sélectionnée pour cela, dans les faits, elle favorise généralement la coordination et la coopération (voire l'altruisme) dans de larges groupes d'individus. Cette hypothèse est en fait très plausible. Dans ce cas, on peut malgré tout attribuer une fonction à la moralité; mais celle-ci doit être comprise en un sens qui relève non des relations historiques et sélectives (version étiologique de la fonction) mais plutôt de relations causales, des effets d'une structure (moralité) sur un système (population d'êtres humains).[406] Cela dit, même si l'on peut attribuer une fonction à la moralité, nous n'en avons pas encore fondé les éléments de base. Il faut ajouter une clause supplémentaire selon laquelle la fonctionnalité de la morale est justifiante, c'est-à-dire que toute valeur ou norme qui favorise la réalisation de la fonction de la moralité (coordination, coopération, voire altruisme) est justifiée.[407]

La dernière manière de fonder les normes et valeurs consiste à se demander si elles ont été sélectionnées au fil de l'évolution grâce aux effets bénéfiques qu'elles apportent ; ainsi, seules les normes et valeurs morales dont on peut montrer que leur application favorise l'adaptation des êtres humains à leur environnement pourraient se justifier (CASEBEER 2003 ; HARMS 2000).

Je pense que lorsqu'il fait appel aux données factuelles (nature humaine, fonction de la morale ou adaptation), le philosophe contemporain se doit de prendre en compte les données scientifiques et évolutionnaires, c'est-à-dire d'adhérer à une forme

[406] La version « propensionniste » par exemple considère la fonction comme une propension d'un système à causer un effet adaptatif (voir PROUST 1995).

[407] Cela implique qu'aucune norme n'est justifiée une fois pour toutes ; si l'environnement change ou si l'évolution prend une tournure nouvelle, une norme peut perdre la caractéristique de remplir la fonction de la moralité (à supposer que la moralité elle-même survit aux changements et garde la même fonction).

d'éthique évolutionniste. Les sections suivantes seront consacrées à une analyse critique du recours aux faits pour justifier les normes et les valeurs morales. Nous verrons que cette approche est l'objet de la fameuse accusation du « passage fallacieux du factuel au normatif ». Il s'agira donc de présenter cette critique avant de considérer diverses manières d'y répondre, c'est-à-dire diverses tentatives d'établir malgré tout un lien entre les données des sciences et le domaine normatif. Face à l'échec d'un certain nombre de ces projets de justification issus de l'éthique évolutionniste, il faudra prendre au sérieux l'idée de l'impossibilité de justifier nos normes et valeurs morales. Je tenterai toutefois de rejeter cette vision défaitiste avant de proposer une position personnelle que j'appellerai la « stratégie du sens commun renforcé ».

7.2. Le lien fallacieux entre le factuel et le normatif ; deux façons d'identifier l'erreur commise

Il est très tentant de chercher à se débarrasser du problème des présupposés injustifiables en choisissant de baser sa théorie morale directement sur des faits scientifiques établis. A priori, l'idée paraît séduisante, mais elle est très controversée. En réalité, ce procédé revient à troquer un problème contre un autre. D'un côté, nous avons trouvé un moyen commode de justifier les éléments de base de la théorie (justification de type scientifique), mais par ailleurs, nous nous retrouvons face à un problème tout autant complexe, celui de la bonne utilisation des faits scientifiques dans le cadre de la pensée morale. Plus précisément, il s'agit de savoir comment, à partir de données scientifiques qui relèvent du domaine des faits, il est possible de dériver des énoncés normatifs. Selon certains philosophes, il s'agit ni plus ni moins d'une mission impossible. D'après Philip KITCHER (1994/1993) par exemple, aucune explication plausible de la manière dont on peut dériver des énoncés factuels à partir d'énoncés normatifs ne peut être proposée. Essayons de comprendre en quoi ce lien entre le factuel et le normatif est fallacieux. Il peut relever de deux sortes d'erreurs : la première est une erreur de type définitionnel et la seconde de type argumentatif.[408]

[408] Dans ce contexte, on parle souvent de « sophisme naturaliste ». Ce terme a été d'abord utilisé par George Edward MOORE (1903) pour désigner l'erreur de type définitionnel. Mais dans les écrits contemporains, il désigne généralement l'erreur de raisonnement thématisée par David HUME (1946/1740, pp. 585-586).

L'erreur de type définitionnel a été dénoncée par George Edward MOORE (1903) et consiste à définir un terme moral (comme « bon » ou « bien ») à l'aide de termes purement descriptifs.[409] Définir complètement le bien en termes factuels est une approche très séduisante car elle permettrait d'offrir une voie royale à la justification : si corrélativement le bien se réduit à des faits, il suffit de connaître ces faits pour connaître le bien. De cette manière, les normes, valeurs et jugements moraux pourraient prétendre à la vérité ou à la fausseté. Toutefois, nous avons vu au chapitre 6 (section 6.1.4) que cette approche réductionniste n'est pas tenable car elle mène à une forme d'éliminativisme. Il faut donc rejeter toutes les tentatives de justification calquées sur ce genre de procédé (ARNHART 1998 ; MCSHEA 1999 ; CASEBEER 2003).

La deuxième façon d'identifier le lien fallacieux entre le normatif et le factuel renvoie à une erreur argumentative qui a été dénoncée par David HUME (1946/1740, pp. 585-586). Elle transgresse ce que l'on appelle généralement la « loi de Hume ». Voyons de quoi il s'agit. Pour justifier un énoncé moral (assertion d'une valeur ou d'une norme, jugement moral) on développe souvent une argumentation logique dans laquelle on pose un certain nombre de prémisses desquelles on déduit l'énoncé moral en question. Dans ce contexte, HUME affirme qu'il n'est pas possible de dériver des conclusions morales uniquement à partir de prémisses purement descriptives (loi de Hume). Cette critique concerne le passage de l'« être » au « devoir-être ». Lisons ce que Hume dit à ce propos :

> « Je ne peux m'empêcher d'ajouter à ces raisonnements une remarque que, sans doute, on peut trouver de quelque importance. Dans tous les systèmes de morale que j'ai rencontrés jusqu'ici, j'ai toujours remarqué que l'auteur procède quelque temps selon la manière ordinaire de raisonner, qu'il établit l'existence de Dieu ou qu'il fait des remarques sur la condition humaine ; puis tout à coup j'ai la surprise de trouver qu'au lieu des copules *est* ou *n'est pas* habituelles dans les propositions, je ne rencontre que des propositions où la liaison est établie par *doit* ou *ne doit pas*. Ce changement est imperceptible ; mais il est pourtant de la plus haute importance. En effet, comme ce *doit* ou ce *ne doit pas* expriment une nouvelle relation et une nouvelle affirmation, il est nécessaire que celles-ci soient expliquées : et qu'en même temps on rende raison de ce

[409] Pour justifier son affirmation, MOORE développe son fameux argument de la question ouverte qui a été présenté à la section 6.1.4 (p. 290). Nous avons vu que cet argument peut être remis en question mais qu'il y a d'autres raisons de rejeter la réduction descriptive contre laquelle MOORE s'est élevé.

> qui paraît tout à fait inconcevable, comment cette nouvelle relation peut se déduire d'autres relations qui en sont entièrement différentes. Mais comme les auteurs n'usent pas couramment de cette précaution, je prendrai la liberté de la recommander aux lecteurs, et je suis persuadé que cette légère attention détruira tous les systèmes courants de morale et nous montrera que la distinction du vice et de la vertu ne se fonde pas uniquement sur les relations des objets et qu'elle n'est pas perçue par la raison. » (HUME 1946/1740, pp. 585-586).

Il est clair que toute éthique évolutionniste se trouve directement confrontée aux objections de MOORE et de HUME, dénonçant le passage fallacieux du factuel au normatif.[410] C'est la raison pour laquelle les défenseurs de cette position consacrent une bonne partie de leur réflexion à échapper à ces deux critiques. Il n'est pas nécessaire de s'attarder sur les propositions de définition et réduction du bien à des entités naturelles puisqu'elles ont déjà été réfutées à la section 6.1.4. La suite de ce chapitre traitera des tentatives d'échapper à la loi de Hume : il sera question de savoir en quel sens il est possible d'affirmer que les données scientifiques jouent un rôle essentiel dans l'élaboration et la justification des normes morales, plus précisément, comment fonder un devoir moral uniquement à partir de données et théories empiriques.

7.3. Les tentatives infructueuses de faire face à la loi de Hume

A première vue, la loi de Hume semble empêcher toute tentative de fonder des énoncés moraux sur des données purement factuelles. Mais beaucoup de défenseurs de l'éthique évolutionniste ont tenté de l'invalider ou de la contourner (pour n'en mentionner que quelques-uns : CASEBEER 2003 ; COLLIER & STING 1993 ; R. CAMPBELL 1996). Parmi ces essais, je présenterai les trois qui me paraissent les plus intéressants. Les deux premiers sont dus à Robert RICHARDS (1986) et le second à Michael RUSE (1998/1986 ; 1993/1991).

[410] On pourrait se demander pourquoi ces arguments ont connu un tel succès en philosophie morale. Dieter BIRNBACHER (1990) en donne une explication assez convaincante. Selon lui, sous-jacent à cette peur du pouvoir des faits, il y a la grande crainte du déterminisme ; peur que l'homme ne soit plus libre de se donner lui-même les règles de conduite qu'il tentera de suivre ; peur aussi que les gens ne se sentent plus responsables de leurs actes. C'est donc au nom de la liberté et de la responsabilité que beaucoup de philosophes affirment sans détour que rien de moral ne peut être déduit à partir des faits.

7.3.1. La stratégie de l'uniformisation de la notion de devoir

Une première manière d'échapper à la loi de Hume serait de dire que la notion de devoir moral ne possède pas un sens tellement différent de celui de devoir empirique, ce dernier étant compris au sens de devoir causal. Robert RICHARDS a développé cette idée en utilisant la notion de « contexte structurant » (1993, p. 129 ; 1986, p. 291). Selon lui, tout devoir émerge dans le cadre d'un contexte. Par exemple, compte tenu des lois physiques de la nature (contexte structurant), si je vois un éclair, je peux me dire qu'un coup de tonnerre *doit* suivre. Considérons un autre exemple : un professeur moyennement exigeant peut dire à un de ses bons étudiants « si tu travailles bien ta matière, tu *dois* passer l'examen » (dans ce cas de figure, le contexte structurant se compose des capacités moyennes des étudiants ainsi que de la quantité et de la difficulté de la matière de l'examen). Dans les deux cas, RICHARDS nous dit que la notion de devoir est de même nature : il s'agit d'un devoir causal (compris au sens de nécessité ou grande probabilité que quelque chose arrive à la suite d'autre chose) dans un contexte précis. Il poursuit cette réflexion dans le domaine moral et affirme que, compte tenu des tendances générales des êtres humains à agir de manière altruiste (contexte structurant), si un jeune homme voit qu'une vieille dame peine à traverser une route dangereuse, dans des circonstances idéales (le jeune homme est normalement constitué, dans un état psychologique et physiologique stable et saisit la détresse de la vieille dame), il *doit* l'aider à traverser la route. En d'autres termes, notre nature nous oblige à agir de manière altruiste lorsque l'occasion se présente.

> « Dans le contexte de la constitution évolutionnaire du comportement humain, le « devoir » [*ought*] signifie qu'une personne doit agir de manière altruiste, à condition qu'il ou elle ait évalué la situation correctement et qu'aucune bouffée de jalousie, haine, avidité, etc. n'interfère. Le « il faut » [*must*] est un « il faut » [*must*] causal ; cela signifie que dans des conditions idéales – c'est-à-dire, des attitudes parfaitement formées résultant de processus évolutionnaires, une connaissance complète des situations, un contrôle absolu sur les passions, etc. – un comportement altruiste sera nécessairement réalisé dans les conditions appropriées. » (RICHARDS 1993, p. 129, ma traduction)

Précisons que pour RICHARDS, les devoirs qui surgissent dans l'exemple de l'éclair, du professeur ou du jeune homme ne sont pas de même nature car ils apparaissent dans des contextes structurants différents ; pour ce qui est de l'exemple du jeune homme, il s'agit d'un devoir *moral* car il apparaît dans un contexte d'action altruiste (RICHARDS 1986, p. 291).

Notons qu'en opérant une telle redéfinition de la notion de devoir, RICHARDS réinterprète les données entrant dans la loi de Hume elle-même, les faisant porter sur un *devoir-être moral de type causal*. Si on accepte cette analyse, alors l'interdiction du passage de l'être au devoir-être moral n'a plus raison d'être puisque les gens utilisent la notion de devoir moral dans le sens d'un simple devoir causal. D'autre part, la seule chose dont nous avons besoin pour définir les normes et valeurs morales est d'acquérir des connaissances empiriques aussi précises que possible au sujet de la nature biologique des êtres humains (plus précisément, leurs dispositions, tendances au comportement).

Une objection qui vient immédiatement à l'esprit est que l'explication de RICHARDS ne parvient pas à rendre compte de notre conviction selon laquelle l'ordre moral est séparé de l'ordre empirique. Plus précisément, elle ne rend pas compte de notre sentiment selon lequel la notion de devoir dans le cas de l'éclair n'est pas du même type que la notion de devoir dans le domaine moral. Après l'éclair, nous ne pouvons pas décider de ne pas entendre le coup de tonnerre (à moins d'avoir un casque antibruit à portée de main…) ; il s'agit d'une nécessité causale. Par contre, même si nous avons le devoir moral de ne pas frapper notre voisin pour le plaisir, il n'en demeure pas moins que nous avons le choix entre frapper notre voisin ou ne pas le faire. Dans son explication, RICHARDS ne rend pas compte de la distinction entre le devoir causal (au sens de nécessité causale ou grande probabilité d'implication causale) et le devoir au sens où nous avons le choix de réaliser ou non une action que le devoir moral prescrit. Pire, ne pas tenir compte de cette distinction revient à nier notre liberté d'action. Or, si l'on admet que la liberté est une condition nécessaire à la responsabilité morale, en adhérant à la position de RICHARDS, nous ne pourrions plus décemment condamner moralement une personne pour une action qu'elle ne pouvait pas éviter d'accomplir ?[411]

[411] A ce propos, voir également les critiques de VOORZANGER 1987 ; P. WILLIAMS 1990 ; LEMOS 1999.

Pour ces raisons, ce que l'on pourrait appeler « la stratégie de l'uniformisation de la notion de devoir » doit être rejetée.[412]

7.3.2. La stratégie de la règle d'inférence

Une autre manière d'échapper à la loi de Hume consiste à élargir l'extension du champ des règles d'inférence. Cet argument est également dû à Robert RICHARDS. Selon lui, il est possible d'inventer de nouvelles règles d'inférence logique, qui permettent de passer d'un certain type de prémisses à un certain type de conclusions. Ainsi, au même titre que les règles courantes des syllogismes,[413] il est possible d'utiliser des règles comme « de 'le leader religieux de la communauté prescrit x' conclure 'les membres de la communauté ont l'obligation morale de faire x' » ou « de 'x est interdit par la bible' conclure 'x est moralement interdit' ». Grâce à ces règles d'inférence, il devient possible de développer des arguments valides contenant uniquement des prémisses descriptives et qui mènent à une conclusion normative (RICHARDS 1993, p. 126). Considérons un exemple proposé par RICHARDS. Imaginons que deux fondamentalistes religieux ne s'accordent pas sur la question de savoir s'il est moralement correct d'avoir des rapports sexuels avant le mariage. Au fil de la discussion, l'un des deux protagonistes parvient à convaincre le second à l'aide de l'argument suivant :

> P_1 : Dans la bible, il est dit que la fornication est interdite (c'est-à-dire : les actes de fornication sont interdits par la bible)
>
> P_2 : Les rapports sexuels avant le mariage sont des actes de fornication
>
> RI_1 : [Si Bx, Ax] → [Si Cx, Bx] → [Si Cx, Ax] (voir note 413)
>
> C_1 : Les rapports sexuels avant le mariage sont interdits par la bible

[412] Il existe peut-être une manière de sauver la position de RICHARDS. On pourrait recourir aux théories de l'illusion de la volonté libre (WEGNER 2002) et dire que, lorsque nous croyons avoir le choix, en réalité, nous ne l'avons pas vraiment ; ce n'est qu'une illusion, tout comme la responsabilité et la morale. Mais c'est un argument très lourd de conséquences et il est douteux que RICHARDS soit prêt à l'admettre.

[413] La plus courante est une règle d'implication qui correspond chez Aristote au premier syllogisme de la première figure. En substance, cette règle dit que si A est affirmé de tout B, et B de tout C, alors nécessairement A est affirmé de tout C. Voici un exemple. Prémisse majeure : Tout oiseau d'eau a des pattes palmées + Prémisse mineure : Tout canard est un oiseau d'eau => Conclusion : Tout canard a des pattes palmées.

RI_2 : [x est interdit par la bible] → [x est moralement interdit]

C_2 : Donc, les rapports sexuels avant le mariage sont moralement interdits[414]

Selon RICHARDS, étant donné que l'argumentation se fonde uniquement sur des prémisses non morales et non impératives (P_1 et P_2), il s'agit d'un exemple de raisonnement valide où l'on passe du factuel (ce qui est le cas) au normatif (ce qui doit être le cas). RICHARDS précise que la conclusion normative C_2 est inférée *au moyen de* la règle d'inférence RI_2 sans pour autant être dérivée *à partir de* cette règle (1986, p. 283). RICHARDS est donc prêt à admettre que dans une communauté de fondamentalistes dont les membres croient à cette règle d'inférence, l'interdiction des rapports sexuels avant le mariage est pleinement justifiée. Cependant, même s'il considère que cet argument est valide et que la conclusion est justifiée dans un certain contexte, RICHARDS ne serait pas prêt à souscrire à cette conclusion car il rejette la règle d'inférence postulée par les fondamentalistes religieux.

Dans le cadre de sa théorie morale, RICHARDS va également utiliser cette idée de règle d'inférence pour fonder des assertions morales ; mais au contraire des fondamentalistes religieux, il cherchera à postuler des règles d'inférence qui correspondent à la manière dont les êtres humains pensent d'ordinaire (car selon lui, la justification des règles d'inférence est une affaire d'acceptation de cette règle par les individus). Celle qu'il propose peut être formulée de la manière suivante : « de 'l'action x promeut le bien de la communauté (c'est-à-dire action x est altruiste)' conclure 'l'action x doit être faite' »[415] (RICHARDS 1989, p. 334). Dès lors, la tâche essentielle à laquelle le lecteur critique se trouve confronté est celle de savoir si la règle d'inférence que RICHARDS propose est acceptable.

La position de RICHARDS peut être critiquée de différentes manières. Premièrement, nous pouvons diriger contre lui une attaque ciblée sur la pertinence de la règle d'inférence qu'il postule. En effet, le seul argument qu'il propose pour soutenir cette règle est un recours au sens commun. Or nous avons vu plus haut (p. 311) à quel

[414] Notons que pour les besoins de l'argument, le raisonnement des fondamentalistes religieux est délibérément simplifié, voire caricaturé par RICHARDS.

[415] « From 'action x promotes the community good' conclude 'x ought to be done' » (RICHARDS 1989, p. 334).

point les justifications de ce type sont limitées.[416] Deuxièmement, la conception que se fait RICHARDS des règles d'inférence est sujette à caution. Tout d'abord, elle ouvre la possibilité d'inventer à peu près n'importe quelle règle d'inférence pour soutenir à peu près n'importe quel énoncé moral (à l'exemple de la règle utilisée dans le raisonnement des fondamentalistes religieux ci-dessus).[417] De plus, en logique, on attend en principe d'une règle d'inférence qu'elle soit entièrement vide de contenu ; ce qui n'est pas le cas des exemples proposés par RICHARDS. Enfin, il faut souligner que l'intérêt de la logique traditionnelle est de mettre en place une procédure fiable qui permet de déduire des conclusions vraies à partir de prémisses vraies. Or si on laisse la place à l'invention de nouvelles règles d'inférence, la logique traditionnelle perd sensiblement de sa force opérationnelle ; on doit non seulement prouver les prémisses mais également les règles d'inférence elles-mêmes ![418] En bref, cette manière de concevoir les règles d'inférence est une grave entorse à la traditionnelle logique des prédicats à laquelle à peu près tous les raisonnements philosophiques se soumettent.

On peut se demander ce que l'on gagne à affaiblir la force opérationnelle de la logique en permettant l'invention de nouvelles règles d'inférence. Aux yeux de RICHARDS, le gain est énorme : ce stratagème lui permet de se débarrasser de la loi de Hume qui semble le hanter (et il n'est pas le seul éthicien évolutionniste dans cette situation). Certes, en invalidant la loi de Hume, RICHARDS peut déduire une conclusion normative uniquement à partir de prémisses factuelles. Mais cela n'empêche pas qu'il n'est pas autorisé à prétendre pouvoir déduire une conclusion normative uniquement à partir de faits, puisque la règle d'inférence en amont de sa conclusion contient précisément la notion de devoir. D'autre part, il n'est pas certain que la loi de Hume soit aussi handicapante pour une éthique évolutionniste que RICHARDS ne veut le croire. Nous verrons plus loin qu'il est tout à fait envisageable de défendre une position morale évolutionniste tout en s'accommodant de la loi de Hume. Pour ces raisons, il paraît déraisonnable de persister à vouloir affaiblir la logique, un outil si central pour les philosophes. En fin de compte, la seule raison d'introduire de nouvelles règles d'inférence semble être celle de donner l'illusion qu'il est possible de contrer la loi de

[416] Notons ici que la stratégie de la règle d'inférence est très peu efficace dans l'entreprise générale de justification des éléments de base de la moralité, puisqu'elle fait appel à des règles d'inférence pour lesquelles il faut encore trouver une justification.

[417] On trouve cette objection notamment chez FERGUSON 2001, pp. 78-79.

[418] Sans mentionner le risque la régression à l'infini.

Hume, laquelle (comme nous le verrons plus loin) n'empêche pourtant pas, par d'autres biais, d'utiliser des données scientifiques et des théories évolutionnistes en philosophie morale.

7.3.3. La stratégie de l'illusion de l'objectivité de la morale

Une autre manière d'établir un lien entre le factuel et le normatif sans s'encombrer de la loi de Hume est la théorie de l'illusion de l'objectivité de la morale qui a notamment été défendue par Michael RUSE (1984 ; 1986 ; 1998/1986 ; 1993/1991), Richard JOYCE (2006) ou Tamler SOMMERS et Alex ROSENBERG (2003).

Considérons dans le détail la réflexion de RUSE. Elle se situe essentiellement au niveau de l'éthique descriptive mais il en tire des conséquences aux niveaux métaéthique et normatif. Brièvement, son argument fonctionne de la manière suivante. La moralité est apparue dans le cadre de l'évolution de l'homme ; elle est une adaptation qui a émergé sous la pression de la sélection naturelle et sa fonction est de rendre sociaux les êtres humains (voir section 5.1, p. 192). D'autre part, puisqu'elle est un simple produit de l'évolution, on ne peut pas lui accorder un statut spécial par rapport à d'autres adaptations biologiques. De ce fait, la morale ne peut pas être fondée ; elle peut seulement être expliquée. Ainsi il écrit : « Une analyse causale du type de celle proposée par les théoriciens évolutionnistes est appropriée et adéquate alors qu'une justification des énoncés moraux en termes de fondements raisonnés n'est ni nécessaire ni appropriée » (RUSE 1986, p. 103, ma traduction).[419]

Si la morale ne peut être fondée, il faut alors expliquer pourquoi les gens pensent pouvoir justifier leurs normes et valeurs. A cet effet, RUSE fait à nouveau appel au mécanisme de la sélection naturelle. Elle aurait façonné la conviction selon laquelle les normes auxquelles nous adhérons sont objectives. Cette croyance en la valeur objective des normes (et valeurs) morales est apparue chez les êtres humains parce qu'elle a pour effet d'inciter les gens à agir de manière coopérative et altruiste, ce qui favorise la

[419] Précisons que si RUSE affirme que la morale n'a pas de fondement ultime (si ce n'est son origine biologique) et ne correspond à aucune réalité indépendante de la subjectivité des sujets, il ne faut pas en déduire qu'il refuse l'existence de la morale. Pour lui la morale existe en tant que phénomène descriptif. Cela lui permet d'affirmer : « Je m'achemine vers ce que l'on appelle souvent le 'scepticisme éthique', en soulignant que le scepticisme porte sur les fondements, non sur les normes » (RUSE 1993/1991, p. 60).

bonne entente et l'élaboration de projets communs (RUSE & E. WILSON 1986, p. 179 ; RUSE 1998/1986, pp. 253-256). Notons que RUSE défend une conception assez complexe de l'objectivité (à ne pas confondre avec l'objectivité de fait et l'objectivité psychologique développées à la section 5.2.1.vi). Il s'agit d'une double croyance causée par nos réactions émotionnelles face à certains types de situations : d'une part la pensée que les normes (et valeurs) doivent être appliquées de manière universelle ; d'autre part la pensée que les normes représentent une réalité morale externe à la subjectivité des sujets.[420]

RUSE ajoute que cette croyance en l'objectivité de la morale n'est qu'une illusion[421] puisqu'elle s'explique simplement par l'action des gènes. Ces derniers nous font réagir émotionnellement face à certaines situations, avec pour conséquence de nous pousser aux comportements coopératifs ou altruistes et de nous insuffler la croyance en l'objectivité de nos convictions coopératives et altruistes (ce qui renforce d'autant plus notre propension à agir en fonction de ces convictions et nous incite à sanctionner les comportements déviants). Ainsi, même si les normes morales ne peuvent pas être justifiées, notre croyance en leur validité s'avère extrêmement utile évolutionnairement parlant.

En bref, au niveau métaéthique RUSE défend une théorie de l'erreur. Il pense que la morale ne possède aucune réalité fixe, universelle, indépendante de la subjectivité des sujets et à laquelle nous aurions un accès cognitif ; et l'idée que les normes morales possèdent une dimension prescriptive et universelle parce qu'elles correspondent à une réalité morale serait une illusion collective produite par les gènes pour nous pousser à agir de manière coopérative.[422] A partir de là, au niveau de l'éthique normative, il en déduit directement l'impossibilité de justifier les énoncés moraux (voir aussi SOMMERS & ROSENBERG 2003).[423]

---

[420] J'expliquerai plus loin (p. 328) pourquoi cette conception de l'objectivité me paraît peu convaincante.

[421] RUSE s'inspire ici de la théorie de l'erreur de John MACKIE (1977) et la complète par une explication étiologique de l'illusion de l'objectivité de la morale.

[422] « La moralité n'a aucun fondement philosophique objectif. Ce n'est qu'une illusion produite en nous [par les gènes] pour promouvoir l''altruisme' biologique. » (RUSE 1986, p. 102, ma traduction)

[423] Mais comme nous le verrons plus loin, il est possible de lui concéder l'antiréalisme sans pour autant accepter l'impossibilité de fonder la morale.

L'avantage de la position de RUSE et de ses successeurs est qu'en admettant uniquement les explications évolutionnaires causales (et non justificatives) de la moralité, elle contourne la loi de Hume :[424] puisque le devoir moral n'est qu'une illusion, ces auteurs ne cherchent pas à déduire un devoir moral à partir de prémisses descriptives (SOMMERS & ROSENBERG 2003, p. 666).

Mais cette théorie comporte un certain nombre de faiblesses. Une objection récente proposée par David LAHTI (2003, p. 644-645) consiste à dire que pour convaincre le lecteur de l'illusion de la moralité, il faut pouvoir montrer l'efficacité évolutionnaire de cette illusion (puisqu'il s'agit d'une adaptation causée par nos gènes). Selon RUSE, l'illusion a pour fonction biologique de nous pousser à coopérer et aider autrui. Mais nous trouvons déjà de la coopération, de l'entraide et de l'altruisme chez les animaux et ces comportements trouvent des explications beaucoup plus classiques (sélection de parentèle, réciprocité, signal coûteux, etc.). Pourquoi imaginer un nouveau mécanisme pour en remplacer d'autres qui fonctionnent déjà très bien ? A cette critique RUSE pourrait rétorquer que les êtres humains se démarquent des animaux par le degré et la qualité de la coopération, de l'entraide et de l'altruisme qu'ils pratiquent. Or, il faut imaginer un nouveau mécanisme pour rendre compte de cette différence ; et ce serait précisément celui de l'illusion de l'objectivité qui serait à la base de la moralité. Toutefois, cette analyse descriptive de la moralité (ou plutôt de l'illusion de la moralité) n'est pas très convaincante dans la mesure où elle la réduit à un objet de sélection. Or nous avons vu au chapitre 5 que de par sa dimension « particulariste », la moralité n'est probablement pas un produit direct de la sélection naturelle (section 5.1, p. 195). L'illusion de l'objectivité de la moralité (comprise à la fois comme croyance en la portée universelle des normes et une croyance en l'existence de propriétés morales extérieures) est quelque chose de bien trop pointu du point de vue du contenu conceptuel pour que l'on puisse raisonnablement penser qu'elle se soit encodée dans

[424] En réalité, RUSE (1993) cherche à utiliser la loi de Hume pour soutenir sa position métaéthique. Puisqu'il n'est pas possible de passer de l'être (niveau factuel) au devoir-être (niveau normatif) et que la moralité est un produit de l'évolution (c'est-à-dire qu'elle se situe au niveau factuel), alors il n'existe aucune nécessité morale (ou normativité) indépendante de la subjectivité des sujets. Mais cet argument est un peu fourvoyant car il mélange différents niveaux. Il est important de comprendre que la loi de Hume concerne une erreur de raisonnement et ne peut pas être utilisée pour tirer des conclusions au niveau ontologique. Il semblerait que RUSE n'est pas au clair sur les différences entre la façon dont MOORE et HUME identifient l'erreur du lien fallacieux entre le factuel et le normatif.

nos gènes (ROTTSCHAEFER & MARTINSEN 1990; LAHTI 2000, p. 646). A cela RUSE pourrait répondre que l'universalité et la prescriptivité ressortent plus du ressenti ou du cognitivement primitif que du réflexif : face à certaines situations, nous ressentons un sentiment de rejet si fort que nous désirons voir les autres partager ce même sentiment et faire en sorte que ces situations ne se reproduisent plus. Ce n'est qu'après coup que nous rationalisons nos réactions émotionnelles et réfléchissons à la portée universelle des normes et valeurs induites par ces réactions, ainsi qu'à leur rapport à une réalité morale extérieure. Certes, mais même si RUSE a raison de dire que les sentiments sont un facteur décisif pour nous faire concevoir l'objectivité (voir section 5.2.1.vi), il n'empêche que la croyance en l'universalité ou en une réalité morale se passe au niveau réflexif. Les émotions n'apportent qu'un input ; l'illusion ne peut donc pas être inscrite dans les gènes.

Au niveau de l'éthique normative, l'objection principale contre la théorie de l'erreur consiste à brandir le spectre du nihilisme. Si tous les sujets moraux prennent conscience de leur illusion (c'est-à-dire du fait que les normes qu'ils édictent ne peuvent pas être fondées, ne possèdent aucune valeur intersubjective et ne correspondent à aucune réalité), alors l'obligation morale et la condamnation morale s'effondrent. Dès lors, nous ne pourrons pas être blâmés d'agir pour notre propre intérêt au détriment d'autrui chaque fois que cela nous arrange. En bref, la position de RUSE et de ses successeurs ouvre la porte au nihilisme moral et à l'épanouissement d'un égoïsme rampant qui, de fait, s'avèrent extrêmement dangereux pour le bon fonctionnement de la société.[425] En conséquence, ne faut-il pas considérer cette théorie morale comme perverse ou du moins biologiquement contre-productive ?

RUSE répond que contrairement à ce que l'on pourrait penser, sa théorie n'est pas liée au nihilisme moral. Les sentiments moraux font partie de la nature humaine ; ils sont le produit d'une adaptation et, à ce titre, sont partagés par l'ensemble des êtres humains. Ainsi, même si nous savons que l'objectivité de la moralité est une illusion, nous ne pouvons pas cesser de la considérer comme objective sous peine de graves troubles psychologiques (RUSE 1998/1986, p. 271). A l'exemple des tourments vécus

[425] On trouve cette critique notamment chez WOOLCOCK 1993 et LEMOS 1999. Notons cependant que cette objection ne porte pas sur la pertinence de la théorie de l'illusion de la moralité elle-même, mais plutôt sur ses conséquences.

par Raskolnikov dans *Crime et Châtiment*,[426] nous ne pouvons pas plus nous passer de la morale que de nos yeux (RUSE 1998/1986, p. 253). Bref, notre tendance naturelle est si forte que, du point de vue de la motivation à l'action, aucun argument philosophique ne peut la contrebalancer.

Tout à fait dans le sens de la réplique de RUSE, Richard JOYCE (2000, pp. 728-729) développe l'argument suivant. Selon lui, il est possible de savoir qu'une règle n'est pas justifiée sans pour autant perdre la motivation à s'y soumettre. Pour éclairer cette idée, JOYCE propose une analogie parlante. Admettons que nous décidions d'entretenir notre condition physique en effectuant un jogging journalier. Sachant que nous avons la fâcheuse tendance à souffrir de faiblesse de volonté en matière d'efforts physiques, nous nous fixons un temps d'effort précis et tâchons de nous y tenir quoi qu'il arrive. Dès lors, nous ressentons fortement l'obligation de suivre la règle « tu dois courir une heure tous les jours ! » alors même que nous savons pertinemment qu'elle ne possède aucune justification profonde (courir dix minutes de moins par-ci par-là ne changera rien à notre condition physique). Sa seule justification est d'ordre pratique (ne pas abandonner le jogging). Selon JOYCE, il en va de même pour les normes morales. Les agents moraux, y compris ceux qui sont convaincus par l'analyse de RUSE (c'est-à-dire convaincus de l'impossibilité de justifier les normes morales), ont conscience de manière plus ou moins confuse que les petits profits à court terme qu'ils pourraient obtenir en enfreignant les normes morales ont des effets désastreux pour la société et sont liés au risque de la punition. Cette prise de conscience confuse, ajoutée à la connaissance de la faiblesse de leur volonté, les pousse à ressentir et s'imposer l'obligation de suivre les ciments de la société que sont les normes morales.[427]

La théorie de la motivation présentée au chapitre 5 (section 5.2.2) semble plutôt parler en faveur de RUSE et JOYCE. Puisque la réflexion théorique n'est qu'indirectement liée à la motivation, il se pourrait bien qu'une prise de conscience de l'illusion de la moralité ne perturbe pas la motivation des agents moraux. Toutefois, j'aimerais montrer qu'il est en réalité assez peu utile de prendre position dans ce débat dans la mesure où l'on peut adhérer dans les grandes lignes à la position métaéthique de

[426] Raskolnikov, le héros du fameux livre de DOSTOÏEVSKY, tue une vieille usurière en se persuadant que sans l'existence de Dieu comme fondement de l'éthique, tout est permis. Après de longs et affreux tourments, Raskolnikov finit par se dénoncer lui-même à la police.

[427] Notons que selon JOYCE, cette motivation n'est pas ordinairement due à un pur calcul rationnel des intérêts propres.

RUSE sans pour autant tirer la conclusion que les normes morales ne peuvent prétendre à aucune forme de justification.

On peut donner raison à RUSE et ses successeurs sur certains points. D'une part la pensée et le comportement moral peuvent être un domaine d'investigation empirique (niveau de l'éthique descriptive) ; d'autre part, comme nous l'avons vu au chapitre 6 (section 6.1), les normes morales ne reflètent pas une réalité immuable et indépendante de la subjectivité des sujets (niveau métaéthique). De plus, RUSE touche un point important en mettant le doigt sur le lien étroit entre nos sentiments (ou réactions émotionnelles) et la croyance en l'objectivité des normes et valeurs que nous défendons. Cependant, comme nous l'avons vu, il n'y a aucune raison d'en faire quelque chose d'inscrit dans nos gènes. D'autre part, il me semble étonnant de faire de cette impression d'objectivité une croyance en la portée *universelle* (et non simplement *intersubjective*) des normes doublée d'une croyance en une réalité morale extérieure (voir p. 326). Même si l'on admet que les gens conçoivent l'objectivité morale dans un sens universaliste (ce que je conteste) et que, du point de vue théorique, l'universalité et le réalisme moral sont deux idées qui s'appellent l'une l'autre, il me paraît tout à fait possible de croire en la portée universelle des normes et valeurs sans pour autant prendre parti sur des questions ontologiques. KANT n'est-il pas l'emblème d'une telle position ? En conséquence, la double conception de l'objectivité (croyance universaliste et réaliste) proposée par RUSE (p. 324) est contestable.

Cela dit, quelle que soit leur position ontologique sur les propriétés morales, on ne peut nier que les êtres humains accordent d'ordinaire une portée intersubjective aux normes et valeurs morales et considèrent que certaines d'entre elles sont plus acceptables que d'autres. Ce faisant, ils ouvrent la porte à la possibilité de justifier les normes et valeurs. Dès lors, toute la question est de savoir comment les justifier.

RUSE et JOYCE postulent une épistémologie morale qui prend les sciences empiriques comme modèle ; une assertion morale est acceptable uniquement dans la mesure où l'on peut prétendre démontrer une correspondance entre celle-ci et son objet (sous entendu que cet objet existe réellement dans le monde). Les théoriciens de l'erreur exploitent ici un idéal que beaucoup de penseurs (notamment RICHARDS ou ROTTSCHAEFER) ont de la peine à abandonner : celui de faire de la morale une théorie scientifique. Ils ont beau jeu ensuite de s'appuyer sur des conclusions métaéthiques

antiréalistes pour montrer que la moralité ne peut être fondée ou justifiée. Mais cette analyse ne tient pas compte d'une réalité fondamentale : le domaine moral touche plus le comportement social que la connaissance et il est plus important de pouvoir convaincre et influencer autrui que d'accéder à une supposée vérité. Dès lors que l'on abandonne l'idéal d'une morale scientifique, l'attrait de la théorie de l'erreur s'évanouit tout simplement. Il y a d'autres moyens que la recherche de la vérité pour convaincre et influencer autrui (voir notamment section 5.2.1.v). La méthode qui intéresse particulièrement les philosophes est celle la justification rationnelle. Dans les sections suivantes, je vais tenter de développer une façon rationnelle de justifier les normes et valeurs morales sans postuler l'existence d'une réalité morale.

7.4. La stratégie du sens commun renforcé

Selon l'approche évolutionnaire, nos croyances morales reflètent certaines de nos dispositions. Celles-ci résultent d'une très longue sélection qui a commencé bien avant l'apparition des hominidés ; elles sont si profondes que la plupart des gens sont incapables de les contrer (même s'il était dans leur intérêt personnel de le faire). Parmi ces dispositions, se trouvent les tendances à la coopération et à l'altruisme. Les théoriciens de l'erreur refusent la possibilité de s'appuyer sur ce genre de dispositions pour fonder les jugements moraux. Cette attitude est bien sûr due au fait qu'ils acceptent pleinement l'impossibilité du passage du factuel au normatif, aussi bien sous sa forme descriptive (MOORE) qu'argumentative (HUME).[428] Quant aux tentatives de passage du factuel au normatif considérées jusqu'à maintenant, le fait qu'elles ne portent pas leurs fruits peut être dû à deux facteurs : d'une part, toutes les tentatives d'établir une correspondance entre nos assertions et une réalité morale indépendante de nos états subjectifs sont vouées à l'échec puisque cette réalité morale n'existe pas (section 6.1) ; d'autre part, les tentatives de fonder la morale uniquement au moyen de raisonnements logiques succombent à la loi de Hume.

[428] Il y a probablement encore d'autres raisons sous-jacentes à ce choix théorique. Par exemple, en acceptant sans précaution le passage du factuel au normatif on peut se retrouver contraint de cautionner des actions généralement considérées comme immorales ; les théories évolutionnistes ne nous montrent-elles pas que nous sommes prédisposés au racisme (FAUCHER & MACHERY 2004 ; BOWLES & CHOI 2004) ?

L'effort de justification des normes et valeurs morales à l'aide de considérations factuelles ne doit pas pour autant être abandonné ; il faut par contre le réorienter, explorer de nouvelles pistes. Dans cette section, j'aimerais montrer qu'il est possible de justifier nos convictions morales tout en acceptant l'antiréalisme, c'est-à-dire sans postuler l'existence de propriétés morales. Cela revient à défendre le parti pris ontologique de RUSE mais à rejeter sa théorie de l'erreur. Mon analyse repose sur l'idée que le domaine moral ne concerne ni la vérité ni la connaissance[429] mais le comportement social des êtres humains. Il faut se souvenir ici de l'approche psychologique qui a été adoptée au chapitre consacré à l'éthique descriptive : s'il y a fondement de nos jugements, cela se passe dans nos esprits ; si quelque chose est considéré comme moralement bon, c'est toujours en fonction de la personne qui émet ce jugement. Dans le cadre de l'activité morale, la justification sert prioritairement à influencer les jugements et l'action des individus. C'est donc en termes de saillance psychologique ou de force de conviction qu'il faut réfléchir. La dimension morale tout comme sa justification ne prennent sens qu'au niveau de ce que pensent les gens ; et il se trouve précisément que les gens sont convaincus de la possibilité de justifier ou garantir une valeur intersubjective à certaines normes morales.

De plus, la notion même de justification doit être conçue de manière assez flexible pour ne pas tomber sous le coup de la loi de Hume. Cette dernière devient assez inoffensive dès que l'on sort du domaine strict de la logique des prédicats ; en effet, tout ce qu'elle dit, c'est qu'un terme (en l'occurrence la composante morale) ne peut pas apparaître dans les conclusions s'il ne figure pas dans les prémisses du raisonnement. Or, du fait qu'une conclusion morale ne peut pas être déduite logiquement à partir de prémisses descriptives[430], on ne peut pas conclure qu'il n'y a *aucune* relation possible entre le descriptif et le moral.[431]

---

[429] En effet, au terme du chapitre 6, il ressort clairement que les sujets moraux n'ont pas d'accès épistémique à une réalité morale extérieure puisque cette dernière n'existe pas.

[430] Il s'agit ici d'une affirmation qui se situe un niveau formel du raisonnement logique ; rien n'est dit à propos de ce qui caractérise les entités contenues dans le raisonnement.

[431] Affirmer cela reviendrait à accepter, sur le plan ontologique, une stricte dichotomie entre le factuel et le normatif. Cette dernière thèse est non seulement hautement sujette à controverse, mais de surcroît, elle ne peut en aucun cas s'appuyer sur un argument purement formel comme la loi de Hume ; on peut assurément souscrire à la loi de Hume sans pour autant défendre une telle dichotomie (à ce propos, voir PUTNAM 2004/2002).

Si la morale n'est pas une affaire de logique, c'est-à-dire qu'elle ne peut être fondée au moyen des outils de la logique formelle, rien n'empêche en revanche d'utiliser d'autres méthodes de justification. Celle que je vais proposer peut être qualifiée de « stratégie du sens commun renforcé ».

7.4.1. Les grandes lignes de la stratégie

Nous avons vu dans la section 7.1 (p. 311) que le recours au sens commun pour fonder un élément de base d'une théorie morale n'est pas très satisfaisant. L'idée que j'aimerais défendre est que si l'on peut soutenir à l'aide de résultats scientifiques et d'arguments théoriques une affirmation attribuable au sens commun, alors on peut dire de cette affirmation qu'elle est justifiée dans la mesure de l'acceptabilité de la théorie et des données qui la soutiennent. En ce sens, on peut parler de recours au « sens commun renforcé » (sous-entendu : renforcé par les données empiriques et les théories évolutionnistes).[432]

Sous-jacente à cette stratégie du sens commun renforcé, il y a l'idée qu'en morale il ne faut pas tant chercher la vérité (comprise au sens de correspondance entre une assertion et la réalité) que des bonnes raisons de croire. James RACHELS (1990, pp. 93-94) défend ce genre de position. Pour lui, nos croyances sont souvent liées entre elles par des connections autres que l'implication logique stricte ; une croyance peut apporter une évidence ou confirmer une autre croyance sans pour autant l'impliquer. Plus les évidences s'accumulent, plus la confiance en une croyance augmente (et inversement). Ainsi, pour produire des raisons de croire à une assertion morale, il n'est pas nécessaire

[432] Cette stratégie m'a été suggérée par les écrits de Robert RICHARDS qui l'avait déjà esquissée en ses propres termes, mais il ne s'est malheureusement pas donné la peine de la mettre en pratique. En effet, le seul argument qu'il propose pour fonder un des principes pivots de sa théorie (à savoir que tous les êtres humains pensent que l'action en faveur du bien de la communauté correspond à l'action morale) est extrêmement faible. Il se contente de préciser que les personnes qui ne sont pas enclines à agir pour le bien de la communauté et pas disposées à penser qu'il s'agit en l'occurrence de la manière morale d'agir, ne correspondent pas au standard humain. Ces personnes seraient des sociopathes qui ont été privés d'un organe nécessaire à l'humanité : le sens moral (RICHARDS 1986, p. 291). Mais cette assertion est hautement critiquable : d'une part il s'agit plus d'une catégorisation que d'un argument, d'autre part, il faudrait admettre dans son sillage un taux incroyablement élevé de sociopathes en ce bas monde… (concernant le deuxième point, voir également JOYCE 2000, pp. 719-720).

d'affirmer que les faits impliquent logiquement l'assertion morale ; il faut plutôt produire les meilleures raisons possibles d'accepter cette assertion. Cette exigence, quoique plus faible, reste significative. Je pense qu'elle est d'autant plus significative lorsqu'elle intègre les convictions du sens commun renforcées par des données scientifiques auxquelles les gens accordent beaucoup de crédit.

Par exemple, si l'on parvient à montrer empiriquement au moyen d'études en anthropologie, psychologie ou économie expérimentale que la quasi-totalité des êtres humains condamne un état de fait (et idéalement que l'on peut proposer une explication étiologique de ce phénomène), alors on peut dire que la condamnation de cet état de fait est justifiée pour les êtres humains, dans l'état actuel de leur évolution. Plus précisément, cette condamnation sera considérée comme justifiée par toutes les personnes qui acceptent les approches scientifiques utilisées. Par exemple, les travaux de Elliott TURIEL et bien d'autres après lui (TURIEL 1983 ; HELWIG & TURIEL 2002 ; NICHOLS 2004) ont montré que dès leur plus jeune âge et dans toutes les cultures, les êtres humains condamnent les actes de violence gratuite envers autrui. Ainsi on dispose d'excellentes raisons de penser que la condamnation de ce type de violence est justifiée (je présenterai d'autres exemples concrets à la section suivante).

Précisons que la stratégie du sens commun renforcé ne succombe pas à la loi de Hume ; il ne s'agit pas de déduire du normatif à partir de données empiriques dans le cadre d'un argument logique, mais plutôt de concevoir une autre manière de justifier un énoncé moral. Ce type de justification ne possède pas une autorité absolue, mais permet de renforcer ou d'affaiblir les assertions morales en fonction des résultats empiriques et de l'élaboration de modèles explicatifs concurrents.

D'autre part, il ne faut pas se méprendre sur la portée de la justification proposée. Le renforcement provenant du domaine scientifique fournit des *raisons de croire* ou des *arguments en faveur de* certains éléments de base. De plus, ces raisons ou arguments sont valables non de manière absolue mais pour des individus qui font l'effort de les comprendre. Je m'éloigne donc de toute éthique évolutionniste qui cherche à fournir des justifications externes, ou valables indépendamment de ce que pensent les sujets moraux (R. CAMPBELL ; RICHARDS ; COLLIER & STINGL ; ARNHART ; CASEBEER 2003 ; etc.). Dans le modèle proposé, la justification a une portée individuelle même si elle procède d'une réflexion rigoureuse : nous devons nous convaincre et convaincre autrui du bien-fondé des normes auxquelles nous adhérons. Ainsi l'éthique normative doit être

conçue comme un processus sans fin de justifications individuelles ; une intense activité intersubjective au cours de laquelle les êtres humains interagissent, échangent et accordent leurs points de vue.[433]

Avant de proposer quelques applications pratiques de la stratégie du sens commun renforcé, j'aimerais préciser un point important. L'exigence de justification au moyen du sens commun renforcé n'est pas propre à la morale et peut être utilisée dans d'autres domaines ; cette stratégie peut tout à fait être appliquée à la justification d'une valeur qui sera ensuite utilisée pour fonder une norme non morale (par exemple d'intérêt rationnel). Ainsi, pour savoir si une norme est pleinement *moralement* justifiée, il faut s'assurer qu'elle remplisse les deux critères de la moralité (recherche de fondement et condition altruiste ; voir section 5.4.1), et que la norme de premier ordre et la/les valeurs sur lesquelles elle s'appuie soient justifiées par le sens commun renforcé. De même, pour savoir si une valeur est pleinement *moralement* justifiée, il faut, outre l'appel au sens commun renforcé, qu'elle remplisse la condition altruiste, c'est-à-dire qu'au cours de l'entreprise de justification, on prenne en considération les intérêts et le bien-être d'autrui.

7.4.2. Justification de quelques éléments de base

Du fait qu'elle repose sur des théories et des connaissances scientifiques d'ordre général, la stratégie du sens commun renforcé ne permet pas de justifier des jugements de valeur ou des normes qui s'appliquent à des contextes particuliers. En principe, on peut uniquement y recourir pour justifier les éléments de base d'un système moral : les valeurs et les normes de premier ordre bien sûr, mais également certains critères de justification qui nécessitent eux-mêmes d'être justifiés.

Voici quelques valeurs qui, à mon avis, peuvent être soutenues par le sens commun renforcé.

Il y a d'abord la valeur de la coopération et la conviction qui lui est associée, selon laquelle toute norme morale valable se doit de favoriser ou du moins ne pas

[433] Allan GIBBARD défend un point de vue similaire dans son excellent livre *Wise Choices, Apt Feelings.*

compromettre la coopération[434] entre les individus.[435] Cette thèse est renforcée par les théories évolutionnistes et un grand nombre de données expérimentales qui ont été détaillées tout au long de cet ouvrage ; nous avons notamment vu qu'une fonction majeure du comportement normatif est de favoriser la coordination et la coopération entre les différents membres d'une société et de combattre l'individualisme excessif (section 2.3.5). Ainsi la validité des normes morales doit être en partie jugée en fonction de leur capacité de promouvoir la coordination et la coopération.

Choisir la valeur de la coopération comme élément de base de la moralité en fait un critère de justification (parmi d'autres) pour des normes morales particulières. Il est intéressant de noter que ce critère implique une certaine relativité morale ; pour juger de la validité d'une norme en termes de sa capacité de promouvoir la coopération, il faut tenir compte des coutumes et croyances propres à la culture dans laquelle cette norme est prônée. Considérons par exemple une communauté qui vit dans une région aride aux ressources limitées et dont les membres pensent que lorsqu'un individu devient socialement inapte (s'il sombre par exemple dans la décadence physique), il se sépare du monde des vivants pour intégrer une zone marginale entre la vie et un monde ancestral accueillant. Une telle croyance était notamment véhiculée dans les dans les années 1860 au sein des communautés Xhosa d'Afrique du Sud (SAGNER 2001 ; voir aussi BROGDEN 2001, pp. 67-68). Dans une telle société, une norme qui incite les membres à abandonner ou précipiter la mort biologique des individus « socialement morts » trouve une forme de justification. En revanche, la même norme serait inadmissible dans une société occidentale contemporaine.

Evidemment la coopération ne peut être qu'une valeur parmi d'autres. Le sens commun renforcé parle également en faveur de la valeur négative de l'inégalité. Les travaux de Christopher BOEHM (1999 ; 2002/2000) montrent que dans la plupart des sociétés humaines, les gens condamnent les inégalités et louent le partage, l'aide et la

[434] On ne peut pas exiger de toute norme morale qu'elle favorise la coopération car certaines normes largement acceptées ne possèdent pas ce genre d'impact (par exemple il n'est pas certain que la prohibition de l'inceste favorise la coopération – du moins pas de manière évidente) ; mais dans ce cas, on peut au moins exiger qu'elles ne compromettent pas la coopération (la norme contre l'inceste réalise cette exigence).

[435] La valeur de la coopération est par ailleurs défendue de manière plus ou moins explicite chez un grand nombre de défenseurs de l'éthique évolutionniste ; comme exemples paradigmatiques, je mentionnerai GIBBARD (2002/1990) ; RICHARDS (1986), RUSE (1998/1986) ou ROTTSCHAEFER (1998/1997).

réciprocité. Il fournit de bonnes raisons évolutionnaires de croire que les impulsions psychologiques sous-jacentes à ces jugements (lesquelles incitent à agir contre l'inégalité) sont de bons moyens de limiter le pouvoir et les abus des individus dominants. Conjointement, nous avons vu que parmi les conditions nécessaires à la stabilité de la cohésion sociale (laquelle est cruciale pour la survie d'un groupe), il y a la pratique de la réciprocité et le comportement d'aide ; or ces derniers sont incompatibles avec l'opportunisme et les abus imposés par des individus dominants (chap. 2). Ainsi, des normes qui prescrivent des traitements inégalitaires (par exemple celle de l'excision) deviennent hautement douteuses ; elles semblent être contraires à notre nature et mettent en péril la stabilité même du système social (cette dernière dépend du pouvoir de contrôler les individus dominants).

De même, la souffrance inutile est une valeur négative qui peut être soutenue comme telle par le sens commun renforcé ; de fait, elle déclenche chez les gens des réactions émotionnelles empathiques très fortes, en particulier lorsqu'elle est causée de manière intentionnelle et lorsqu'elle est produite par un contact physique direct (section 5.2.1.iii, p. 200).

Le sens commun renforcé permet également de soutenir des critères de justification ou de pondération d'autres éléments de base (en l'occurrence les valeurs ou les normes de premier ordre). Par exemple, l'exigence de cohérence est une réalité empirique, une force interne ressentie par tout être rationnel qui le pousse à reconsidérer les normes ou les valeurs qui entrent en contradiction (voir section 5.2.1.v, p. 211). Etant donné cette réalité, on peut soutenir la valeur épistémique (ou critère de justification) de la cohérence. De même, étant donné que la moralité ressort du domaine pratique, il faut admettre une certaine dose de pragmatisme dans la théorie. Ainsi, les normes doivent être suffisamment efficaces pour remplir le rôle principal de la moralité qui consiste à renforcer la cohésion sociale.[436] Ou alors, la portée des normes doit être relative à la distance (ce que B. WILLIAMS appelait *relativism of distance* ; 1985) ; cette

[436] Cela implique par exemple une hiérarchisation des normes qui favorise celles qui garantissent les besoins les plus fondamentaux des êtres humains. Pour définir cette hiérarchie, on peut faire appel à des études empiriques. Par exemple, sur la base d'observations réalisées dans les années 1940, le psychologue Abraham MASLOW (1943) a élaboré un modèle hiérarchique des besoins humains, avec, par ordre d'importance, les besoins physiologiques, la sécurité, l'amour et l'appartenance, l'estime des autres, l'estime de soi et enfin l'accomplissement personnel.

dernière peut être comprise en termes de relations sociales, de degré de parenté ou de distance spatiale (voir aussi RUSE 1993, p. 57).[437]

Aucun de ces différents éléments cautionnés par le sens commun renforcé ne paraît suffisant en lui-même pour fonder l'ensemble d'un système moral.[438] Par exemple, toute norme efficace ou qui promeut la coopération ne peut pas forcément être considérée comme morale ; tout système de normes cohérent n'est pas forcément moralement acceptable, etc. C'est à chaque être humain de prendre en considération les différents paramètres des situations concrètes auxquelles il est confronté et de les évaluer en fonction de ce qu'il ressent et des valeurs, normes et critères auxquels il adhère.

7.4.3. Réponse à deux critiques : le cercle méthodologique et le scepticisme

L'éthique évolutionniste est souvent accusée de commettre l'erreur du cercle méthodologique. Elle postule que la coopération est un bien ou que l'inégalité est un mal, puis en donne une explication évolutionnaire censée être à l'origine de ce bien ; mais cette explication n'est pas créatrice de bien ou de mal, elle ne dit pas en quoi la

[437] Il existe une large littérature évolutionnaire sur les vertus adaptatives des comportements d'aide prodigués de manière discriminante (voir notamment section 2.2.1.iv ; pour une revue de cette littérature, voir aussi JOYCE 2006, pp. 46-47).

[438] A cet égard, certains éthiciens évolutionnistes sont trop optimistes. John COLLIER et Michael STINGL par exemple pensent qu'en réfléchissant sur les conditions générales dans lesquelles l'évolution agit et sur la meilleure manière dont des créatures sociales hautement évoluées et intelligentes peuvent survivre dans l'environnement que l'on connaît, il doit être possible de découvrir les principes moraux que ces créatures devraient raisonnablement suivre. « Knowledge of evolutionary theory and adaptive processes allows us to speculate about what our moral instincts might have been had our morality evolved more optimally. This, in turn, allows us to formulate (still as a process of empirical discovery) what general moral principles might apply to optimally evolved, intelligent, social creatures. » (1993, pp. 55-56)
Cette solution pèche par son idéalisme. Indépendamment du fait qu'ils admettent implicitement, et sans argument supplémentaire, que la survie de l'espèce ou le bien-être commun sont bons en soi, si ces auteurs menaient leur raisonnement jusqu'au bout, ils comprendraient que l'optimum évolutionnaire n'existe pas (voir à ce propos DE SOUSA 2007). Il est important de comprendre que les données empiriques et théories évolutionnistes ne découvrent aucune valeur ou norme ; elles fournissent uniquement des raisons d'accorder notre confiance à certaines d'entre elles (étant donné le monde dans lequel nous vivons, nos biais psychologiques, etc.).

coopération ou l'inégalité sont particulières par rapport à d'autres produits de l'évolution. Ainsi, Philippe HUNEMAN écrit : « L'éthique évolutionniste *importe* des jugements moraux, qui sont précisément toute la partie éthique de son discours (le reste étant une contribution à l''histoire naturelle de la morale'), avant de les proférer de nouveau comme conséquences de façon assez circulaire » (2007, p. 246). Je pense que cette critique peut être aisément désamorcée si l'on ne réduit pas l'éthique évolutionniste à l'éthique descriptive, et si l'on prend conscience du fait que *toute* théorie morale est contrainte, à un moment donné, de poser un certain nombre de prémisses a priori (ce que j'ai appelé les « éléments de base »). Cela étant admis, il me semble que la réflexion dirigée à la fois par l'intuition du sens commun et par les théories évolutionnistes et données empiriques nous fournit les meilleures raisons possibles de postuler certains éléments de base plutôt que d'autres. Il y a donc des *conditions particulières* dans lesquelles il est possible d'établir un lien de justification entre le factuel et le normatif.

Beaucoup de philosophes de la morale pensent que les normes morales possèdent une valeur universelle. Au contraire, mon analyse inspirée des modèles de coévolution gène-culture, indique qu'il est vain de chercher à découvrir un code moral universel. La validité des normes est dépendante à la fois de la nature humaine (en tant qu'espèce biologique produite par l'évolution) et de la culture de la société dans laquelle elles sont énoncées. Ainsi, il est nécessaire que les éléments de base de la théorie morale soient suffisamment flexibles pour que les normes qui en découlent puissent s'adapter aux circonstances sans cesse fluctuantes de l'évolution culturelle et génétique. De plus, un ensemble de normes et valeurs est toujours bon pour des individus en fonction du contexte dans lequel il prend place.[439] Nous trouvons les conditions de justification de nos jugements, normes et valeurs au moyen de notre propre ressenti et de notre raisonnement ; et non par rapport à une référence externe (Dieu, Idée platonicienne, raison pure, nature humaine conçue comme quelque chose de fixe, etc.).[440]

On pourrait craindre qu'un tel système mène au scepticisme ou au relativisme absolu. Cependant, à l'inverse de certains auteurs (QUINE 1986), je ne pense pas que le

[439] On retrouve ici, mais avec une teinte évolutionnaire, la thèse de Bernard WILLIAMS (1985) selon laquelle les raisons d'agir doivent provenir de la subjectivité des agents eux-mêmes ; elles doivent être internes plutôt qu'externes. L'éthique est une affaire de perspective et échappe à toute universalité.

[440] A ce propos, voir aussi les écrits de John DEWEY (1994).

caractère ouvert et indéterminé du champ des valeurs et normes morales nous impose un tel scepticisme (ni même un relativisme radical). L'évolution a forgé un certain nombre de constantes qui influencent nos réflexions et choix moraux. Parmi ceux-ci, il y a les besoins fondamentaux des êtres humains, leurs biais psychologiques (de contenu et de transmission), ou le fait qu'un facteur nécessaire à la stabilisation d'un comportement social consiste en ce qu'il favorise la coordination et la coopération à l'intérieur d'une société. Tous ces facteurs associés à un effort de communication garantissent un certain degré d'accord sur des convictions morales saines.

Conclusion

L'objectif de ce chapitre était de montrer qu'il est possible de mener un projet d'éthique évolutionniste au niveau normatif. Toutefois, il est clair que cela ne peut se faire sans précaution. A l'examen des critiques de MOORE et de HUME, nous avons vu que les données et théories scientifiques ne fournissent pas les moyens de définir (et réduire) le bien moral, pas plus qu'elles ne permettent de justifier des assertions morales de manière définitive (au sens de l'implication logique stricte) ou de définir une quelconque vérité morale stable. Sur la base de ces considérations, Michael RUSE a tiré la conclusion que les êtres humains vivent dans l'illusion de l'objectivité des normes morales. Cette solution ne me paraissant pas convaincante, j'ai proposé, partant des mêmes leçons philosophiques (MOORE, HUME) et évolutionnaires, un modèle de justification morale particulariste qui applique la stratégie du sens commun renforcé. Cette solution repose sur une nouvelle conception de l'entreprise même de la justification des normes et valeurs morales ; sur la base du constat que l'éthique ne pourra jamais prétendre à l'exactitude scientifique, j'ai proposé de faire l'économie des notions de vérité et d'universalité morale et de s'efforcer de développer les meilleures justifications individuelles possibles, celles qui nous permettent de convaincre autrui de la pertinence de nos assertions morales.

Je ne prétends pas fonder les normes et valeurs morales sur la seule base de considérations empiriques et évolutionnaires. Cette entreprise nécessite, de la part de chaque agent moral, le recours à une réflexion d'ordre philosophique qui lui permet de fonder ses jugements moraux sur la base de valeurs et de normes, lesquelles trouvent une forme de justification au moyen de l'application de la stratégie du sens commun

renforcé. C'est au moyen de cette entreprise de justification des éléments de base du système moral que l'éthique évolutionniste prend son sens au niveau normatif. Elle seule peut fournir une justification aux éléments de base d'un système moral ; mais cette justification n'est que de type contextuel et non absolu. En fin de compte, le domaine de l'éthique normative apparaît comme un processus en constante évolution qui se calque sur le ressenti et le bon sens des gens et s'adapte aux circonstances nouvelles.

Du point de vue pratique, le sens commun renforcé donnera des résultats appréciables uniquement si les êtres humains mettent leurs efforts en commun pour s'entendre sur ce que dit leur sens commun, à savoir s'ils sont disposés à se remettre en question, à confronter leurs convictions avec celles des autres et avec les données des sciences. Comme le dit Allan GIBBARD (2002/1990), ce n'est que par le bais de la discussion normative (laquelle est guidée par différents mécanismes psychologiques) à l'intérieur et entre les communautés que nous parvenons à nous entendre sur les exigences de la moralité. J'ajouterai qu'en faisant l'effort de fonder l'ensemble des éléments de base de nos systèmes moraux sur le sens commun renforcé par des théories et données empiriques fiables, il doit être possible d'atteindre l'objectif le plus recherché en philosophie morale : développer des systèmes moraux robustes, crédibles et adaptés aux êtres humains.[441] Ce projet dépend cependant de notre foi en la science ainsi que des progrès de l'analyse descriptive du comportement et de la pensée morales (les mécanismes psychologiques sur lesquels reposent ces phénomènes, leur fonction évolutionnaire, etc.).

[441] Ainsi, Jeffrie MURPHY n'a que partiellement raison lorsqu'il écrit : « Nous rejetons l'utilitarisme simpliste parce qu'il implique des conséquences moralement contre-intuitives, ou nous adhérons à une théorie de la justice rawlsienne parce qu'elle systématise (au moyen de 'l'équilibre réfléchi') nos convictions préthéoriques. Mais quel est le statut de nos intuitions ou de nos convictions ? Peut-être n'y a-t-il rien de plus à en dire que le fait qu'elles impliquent des préférences (ou des systèmes de préférences) profondément ancrées dans notre nature biologique. » (MURPHY 1982, p. 112, n.21) Je pense que grâce à la stratégie du sens commun renforcé, nous pouvons précisément dire quelque chose de plus : en disposant d'une explication des préférences profondément ancrées en nous, nous pouvons accorder une certaine légitimité à nos intuitions et convictions (et inversement si aucune justification empirique ne peut être fournie), et, partant, à certains systèmes philosophiques plutôt qu'à d'autres.

Conclusion

Au cours de l'histoire de la philosophie, l'éthique et en particulier la problématique de la justification morale semblaient avoir gagné de l'autonomie par rapport aux sciences. L'éthique évolutionniste est l'acteur d'une inversion de ce courant ; aujourd'hui, les sciences reprennent indéniablement du terrain dans ce domaine de pensée.

Le présent ouvrage avait pour objectif d'analyser la manière dont les données et théories scientifiques peuvent être intégrées dans une réflexion éthique ; il s'agissait de sonder les limites et les possibilités d'une éthique évolutionniste. De cette analyse, il ressort que la moralité est une production humaine ; il s'agit d'une construction de nos esprits qui nous aide à vivre en société. Son domaine d'application peut uniquement être délimité d'un point de vue théorique. J'ai proposé deux critères d'individuation qui me paraissent crédibles et semblent refléter le champ d'application que l'on attribue couramment à la morale.

Une autre thématique importante traitée dans cet ouvrage concerne l'influence de notre passé évolutionnaire sur la manière dont nous évaluons les situations sociales, dont nous formons nos assertions morales et sur ce qui nous motive à agir. J'ai montré comment cette activité évaluative repose sur des capacités, biais psychologiques et mécanismes émotionnels sélectionnés au cours de l'évolution en réponse à diverses pressions du monde dans lequel nos ancêtres ont vécu. Une problématique qui mériterait d'être approfondie est celle de l'impact de la réflexion sur nos évaluations spontanées. Dans quelle mesure permet-elle de transcender les déterminants génétiques (mécanismes et biais psychologiques) et culturels ? Les conclusions auxquelles je suis arrivée dans ce livre laissent cependant présager une portée relativement faible de nos raisonnements sur nos évaluations spontanées.

La moralité a toujours été entourée d'une aura qui lui confère une certaine noblesse. Cela provient probablement du fait qu'elle nous a longtemps échappé ; il n'est pas évident de comprendre pourquoi certaines situations déclenchent en nous des réactions d'approbation ou de désapprobation aussi violentes. L'éthique évolutionniste nous en fournit une première explication. On pourrait objecter qu'elle repose sur des hypothèses de portée considérable, en commençant par la théorie de l'évolution elle-

même. Certes, mais en sciences la crédibilité des hypothèses avancées est mesurée à leur force explicative. A cet égard, il me semble avoir pu montrer dans cet ouvrage que l'éthique évolutionniste a tenu son pari. Nous pouvons nous en réjouir quoiqu'il faille admettre un revers de la médaille à ces avancées théoriques : à mesure que l'on s'achemine vers une meilleure compréhension du fonctionnement de la pensée et de l'activité évaluative, la morale perd de son mystère. Il y a quelque chose de désagréable et de déstabilisant dans cette démystification. Faut-il dès lors regretter l'avènement de l'éthique évolutionniste ? Je ne le pense pas. Il fut un temps où dans l'opinion commune, l'univers et ses planètes représentaient ou étaient régis par des divinités ; toutes sortes de mythes rendaient compte de la forme et du mouvement des astres qui composent notre univers. Après les grandes découvertes astronomiques, ces interprétations se retrouvent consignées dans de passionnantes compilations de récits mythologiques… Je ne vois aucune raison de s'en plaindre. De plus, pour revenir à la morale, une compréhension détaillée de son fonctionnement ne nous empêchera pas de nous indigner devant certaines situations et de considérer nos réactions comme devant être intersubjectivement partagées ; la morale ne pourra probablement jamais être démystifiée du point de vue de la psychologie individuelle.

Indépendamment de ses effets de sur l'aura de la moralité, l'éthique évolutionniste n'a pas sonné le glas de la philosophie morale. En réalité, elle s'y intègre harmonieusement : plus qu'un courant philosophique à proprement parler, l'éthique évolutionniste est l'expression d'une nouvelle méthodologie appliquée en philosophie morale.

Dans le cadre même de l'éthique évolutionniste, le présent ouvrage a son intérêt puisque je prends position dans les débats qui font rage dans ce courant de pensée. Au niveau descriptif je rejette les explications de la moralité comme adaptation ; au niveau métaéthique j'argumente contre toute forme de réalisme moral et défends une position antiréaliste particulière ; au niveau normatif, je donne raison à HUME et MOORE sur la question du passage fallacieux du factuel au normatif et explore d'autres possibilités de justifier nos jugements moraux à l'aide de données factuelles. Enfin, la plus grande originalité de cet ouvrage tient sans doute dans la décomposition de l'altruisme psychologique en deux formes, l'une motivationnelle et l'autre sophistiquée. Cette procédure dénoue différents problèmes. Elle autorise à sonner le glas de la thèse de l'égoïsme psychologique. Elle permet de préciser les liens entre l'altruisme

évolutionnaire et l'altruisme psychologique : la forme motivationnelle du second repose sur l'évolution du premier. Enfin, elle éclaire le rapport entre l'altruisme psychologique et la morale : l'altruisme motivationnel est une forme de motivation morale alors que l'altruisme sophistiqué constitue un critère d'individuation de la moralité.

Bibliographie

ACTON, Harry, 1936, "The Expletive Theory of Morals", *Analysis*, 4, pp. 42-45.

AGAR, Nicholas, 2002, "Agar's Review of Katz", *Biology and Philosophy*, 17, pp. 123-139.

ALEXANDER, Richard, 1979, *Darwinism and Human Affairs*, Washington: University of Washington Press.

ALEXANDER, Richard, 1987, *The Biology of Moral Systems*, Hawthorne: Aldine de Gruyter.

ALEXANDER, Richard, 1993, "Biological Considerations in the Analysis of Morality", in M. NITECKI *et al.* (éds.), *Evolutionary Ethics*, Albany: SUNY Press, pp. 163-196.

ALVARD, Michael, 2003, "The Adaptive Nature of Culture", *Evolutionary Anthropology*, 12, pp. 136-149.

ANSCOMBE, Elisabeth, 1958, "Modern Moral Philosophy", *Philosophy 33*, 124 ; in W. HUDSON *et al.* (éds.), 1973 (1969), *The Is/Ought Question; A collection of Papers on the Central Problem in Moral Philosophy*, New York: St.Martin's Press (Controversies in Philosophy), pp. 175-195.

ARISTOTE, 1992 (4ème s. av. J.-C.), *Ethique à Nicomaque*, Paris: Livre de Poche (Classiques de la philosophie).

ARNHART, Larry, 1998, *Darwinian Natural Right; The Biological Ethics of Human Nature*, New York: State University of New York Press.

ARNHART, Larry, 2000, "The Search for a Darwinian Science of Ethics", *Dialogos* (en ligne), 13, pp. 1-7. <http://dialogos3.tripod.com/index.htm> (réf. 20.12.2005)

ATRAN, Scott, 2001, "The Trouble with Memes; Inference versus Imitation in Cultural Creation", *Human Nature*, 12/4, pp. 351-381.

AXELROD, Robert, 1986, "An Evolutionary Approach to Norms", *American Political Science Review*, 80, pp. 1095-1111.

AXELROD, Robert, 1996 (1984), *Comment réussir dans un monde d'égoïstes ?*, trad. de l'angl. par M. GARÈNE, Paris: Odile Jacob.

AXELROD, Robert, HAMILTON, William, 1996 (1984), "L'évolution de la coopération dans les systèmes biologiques", in R. AXELROD, 1996 (1984), *Comment réussir dans un monde d'égoïstes ?*, trad. de l'angl. par M. GARÈNE, Paris: Odile Jacob, pp.87-102.

AYER, Alfred, 1946 (1936), *Language, Truth, and Logic*, New York: Dover Publications.

BALDWIN, James, 1896, "A New Factor in Evolution", *American Naturalist*, 30, pp. 441-451, pp. 536-553.

BALDWIN, James, 1980 (1909), *Darwin and the Humanities*, New York: AMS Press.

BARKOW, Jerome, 1989, *Darwin, Sex, and Status; Biological Approaches to Mind and Culture*, Toronto: University of Toronto Press.

BARON, Jonathan, 1998, *Judgment Misguided; Intuition and Error in Public Decision Making*, Oxford: Oxford University Press.

BARRETT, Louise, HENZI, Peter, WEINGRILL, Tony, LYCELL, John, HILL, Russell, 1999, "Market Forces Predict Grooming Reciprocity in Female Baboons", *Proceedings of the Royal Society of London B*, 266, pp. 665-670.

BARTH, Jochen, POVINELLI, Daniel, CANT, John, 2004, "Bodily Origins of SELF ", *in* D. BEIKE *et al.* (éds.), *The Self and Memory*, New York: Psychology Press, pp. 11-43.

BATSON, Daniel, 1991, *The Altruism Question; Toward a Social-Psychological Answer*, Hillandale: Lawrence Earlbaum.

BEER, Jennifer, HEEREY, Erin, KELTNER, Dacher, SCABINI, Donatella, KNIGHT, Robert, 2003, "The Regulatory Function of Self-Conscious Emotion; Insights from Patients with Orbitofrontal Damage", *Journal of Personality and Social Psychology*, 85/4, pp. 594-604.

BENTHAM, Jeremy, 1948 (1789), *An Introduction to the Principles of Morals and Legislation*, Oxford: Blackwell.

BEN-ZE'EV, Aaron, 2000, *The Subtlety of Emotions*, Cambridge (MA): MIT Press.

BIRNBACHER, Dieter, 1990, "Rechte des Menschen oder Rechte der Natur? Die Stellung der Freiheit in der ökologischen Ethik", *Studia Philosophica*, 40, pp. 61-79.

BLACKBURN, Simon, 1993, *Essays in Quasi-Realism*, Oxford: Oxford University Press.
BLACKBURN, Simon, 2000 (1998), *Ruling Passions; A Theory of Practical Reasoning*, Oxford: Clarendon Press.
BLACKMORE, Susan, 1999, *The Meme Machine*, Oxford: Oxford University Press.
BLAIR, James, MITCHELL, Derek, BLAIR, Karina, 2005, *The Psychopath: Emotion and the Brain*, Oxford: Blackwell.
BLOCK, Ned, 1995, "On a Confusion about a Function of Consciousness", *Behavioral and Brain Sciences*, 18, pp. 227-287.
BLUMSTEIN, Daniel, STEINMETZ, Jeff, ARMITAGE, Kenneth, DANIEL, Janice, 1997, "Alarm Calling in Yellow-Bellied Marmots; II. The Importance of Direct Fitness", *Animal Behaviour*, 53, pp. 173-184.
BOEHM, Christopher, 1997, "Impact of The Human Egalitarian Syndrome On Darwinian Selection Mechanics", *The American Naturalist*, 150, pp. 100-121.
BOEHM, Christopher, 1999, *Hierarchy in the Forest; The Evolution of Egalitarian Behavior*, Cambridge (MA): Harvard University Press.
BOEHM, Christopher, 2002 (2000), "Conflict and the Evolution of Social Control", in L. KATZ (éd.), 2002 (2000), *Evolutionary Origins of Morality;Cross-Disciplinary Perspectives*, Bowling Green: Imprint Academic, pp. 79-101.
BOESCH, Christophe, 1996, "The Emergence of Cultures among Wild Chimpanzees", *Proceedings of the British Academy*, 88, pp. 251-268.
BOESCH, Christophe, TOMASELLO, Michael, 1998, "Chimpanzee and Human Culture", *Current Anthropology*, 39, pp. 591-604.
BOREL, Emile, 1921, "La théorie des jeux et les équations intégrales à noyau symétriques", *Comptes Rendus de l'Académie des Sciences*, 173, pp. 1304–1308.
BOUCHARD, Thomas, 2004, "Genetic Influence on Human Psychological Traits", *American Psychological Society*, 13, 4, pp. 148-151.
BOURKE, Andrew & FRANKS, Nigel, 1995, *Social Evolution in Ants*, Princeton, New York: Princeton University Press.
BOYD, Richard, 1988, "How to be a Moral Realist", in G. SAYRE-MCCORD (éd.), *Essays on Moral Realism*, Ithaca, London: Cornell University Press, pp. 187-228.
BOYD, Robert, GINTIS, Herbert, BOWLES, Samuel, RICHERSON, Peter, 2003, "The Evolution of Altruistic Punishment", *Proceedings of the National Academy of Sciences of the United States of America*, 100, pp. 3531-3535.
BOYD, Robert, LORBERBAUM, Jeffrey, 1987, "No Pure Strategy is Evolutionary Stable in the Repeated Prisoner's Dilemma Game", *Nature*, 327, pp. 58-59.
BOYD, Robert, RICHERSON, Peter, 1985, *Culture and the Evolutionary Process*, Chicago: University of Chicago Press.
BOYD, Robert, RICHERSON, Peter, 1988, "The Evolution of Reciprocity in Sizable Groups", *Journal of Theoretical Biology*, 132, pp. 337-356.
BOYD, Robert, RICHERSON, Peter, 1992, "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups", *Ethology and Sociobiology*, 13, pp. 171–195.
BOYD, Robert, RICHERSON, Peter, 1996, "Why Culture is Common, but Cultural Evolution is Rare", *Proceedings of the British Academy*, 88, pp. 77-93.
BOWLES, Samuel, CHOI, Jung-Kyoo, 2004, "The Co-evolution of Love and Hate", in P. VAN PARIJS (éd.), *Cultural Diversity versus Economic Solidarity*, Brussels: De boeck, pp. 189-202.
BOWLES, Samuel, FEHR, Ernst, GINTIS, Herbert, 2003, "Strong Reciprocity May Evolve With or Without Group Selection", *Working Papers of the Institute for Empirical Research in Economics in Zürich* (en ligne). <www.iew.unizh/home/fehr> (réf. 13.07.2005)
BOWLES, Samuel, GINTIS, Herbert, 2002, "Prosocial Emotions", *Santa Fe Working Paper* (en ligne). <http://www.santafe.edu/research/publications/working-papers.php> (réf. 06.09.2006)
BOWLES, Samuel, GINTIS, Herbert, 2004, "The Evolution of Strong Reciprocity; Cooperation in Heterogeneous Populations", *Theoretical Population Biology*, 65, pp. 17-28.

BRADIE, Michael, 1994, *The Secret Chain; Evolution and Ethics*, New York: State University of New York Press (SUNY series in philosophy & biology).
BRANDT, Hannelore, HAUERT, Christoph, SIGMUND, Karl, 2006, "Punishing and Abstaining for Public Goods", *Proceedings of the National Academy of Sciences of the United States of America*, 103/2, pp. 495-497.
BRANDT, Richard, 1959, *Ethical Theory*, Englewood Cliffs: Prentice-Hall.
BRINK, David, 1989, *Moral Realism and The Foundations of Ethics*, Cambridge, New York: Cambridge University Press.
BROAD, Charlie, 1930, *Five Types of Ethical Theory*, London: Routledge & Kegan Paul.
BROAD, Charlie, 1971/1953, "Self and others", in H. LEWIS (éd.), 1971, *Broad's Critical Essays in Moral Philosophy*, London: Allen & Unwin, pp. 277–282.
BROGDEN, Mike, 2001, *Geronticide; Killing the Elderly*, London, Philadelphia: Jessica Kingsley Publishers.
BROSNAN, Sarah, DE WAAL, Frans, 2003, "Monkeys Reject Unequal Pay", *Nature*, 425, pp. 297-299.
BROWN, Donald, 1991, *Human Universals*, New-York: McGraw-Hill.
BSHARY, Redouan, SCHÄFFER, Daniel, 2002, "Choosy Reef Fish Select Cleaner Fish that Provide High-Quality Service", *Animal Behaviour*, 63, pp. 557-564.
BUTLER, Joseph, 1991/1726, *Fifteen Sermons*, in D. RAPHAEL (ed), 1991 (1969), *British Moralists*. Vol. 1, Indianapolis, Cambridge: Hackett Publishing Company, pp. 325-377.
CABANAC, Michel, GUILLAUME, Jacqueline, BALASKO, Marta, FLEURY, Adriana, 2002, "Pleasure in Decision-Making Situations", *BMC Psychiatry* (en ligne), 2/7. <http://www.biomedcentral.com/1471-244X/2/7> (réf. 15.05.2007)
CAMPBELL, Neil, REECE, Jane, 2004 (1995), *Biologie*, trad. de l'angl., adapt. et rév. scient. par R. MATHIEU, Bruxelles: De Boeck Université.
CAMPBELL, Richmond, 1996, "Can Biology make Ethics objective?", *Biology and Philosophy*, 11, pp. 21-31.
CASEBEER, William, 2003, *Natural Ethical Facts; Evolution, Connectionism, and Moral Cognition*, Cambridge: MIT Press.
CASHDAN, Elizabeth, 1989, "Hunters and Gatherers; Economic Behavior in Bands", in S. PLATTNER (éd.), *Economic Anthropology*, Stanford: Stanford University Press, pp. 21-48.
CAVALLI-SFORZA, Luigi, FELDMAN, Marcus, 1981, *Cultural Transmission and Evolution; A Quantitative Approach*, Princeton, New York: Princeton University Press.
CHAPUISAT, Michel, KELLER, Laurent, 2007, "Les Fourmis, en froid avec Darwin?", *Les dossiers de la recherche*, 27, pp. 56-63.
CHEN, Serena, SHECHTER, David, CHAIKEN, Shelly, 1996, "Getting at the Truth or Getting Along; Accuracy- versus Impressionmotivated Heuristic and Systematic Processing", *Journal of Personality and Social Psychology*, 71/2, pp. 262-275.
CHENEY, Dorothy, SEYFARTH, Robert, 1990, *How Monkeys see the World*, Chicago: University of Chicago Press.
CHOMSKY, Noam, 1965, *Aspects of the Theory of Syntax*, Cambridge (MA): MIT Press.
CHURCHLAND, Paul, 1998, "Toward a Cognitive Neurobiology of the Moral Virtues", *Topoi*, 17, pp. 83-96.
CIALDINI, Robert, SCHALLER, Mark, HOULIHAN, Donald, ARPS, Kevin, FULTZ, Jim, BEAMAN, Arthur, 1987, "Empathy-Based Helping; Is It Selflessly or Selfishly Motivated? ", *Journal of Personality and Social Psychology*, 52, pp. 749-758.
CIARAMELLI, Elisa, MUCCIOLI, Michela, LÀDAVAS, Elisabetta, DI PELLEGRINO, Giuseppe, 2007, "Selective Deficit in Personal Moral Judgment following Damage to Ventromedial Prefrontal Cortex", *SCAN*, 2, pp. 84-92.
CLAIDIÈRE, Nicolas, SPERBER, Dan, 2007, "The Role of Attraction in Cultural Evolution", *Journal of Cognition and Culture*, 7, pp. 89-111.
CLAVIEN, Christine, soumis, "An Affective Picture of Values and Moral Judgements"

CLEMENT, Fabrice, 2007, "Du proto-soi social au sujet moral ; rupture ou continuité?", in C. CLAVIEN *et al.* (éds.), *Morale et évolution biologique ; entre déterminisme et liberté*, Lausanne: PPUR, pp. 170-190.
COLLIER, John, STINGL, Michael, 1993, "Evolutionary Naturalism and the Objectivity of Morality", *Biology and Philosophy*, 8, pp. 47-60.
CONNOR, Richard, 1995, "Impala Allogrooming and the Parceling Model of Reciprocity", *Animal Behaviour*, 49, pp. 528-530.
COPP, David, ZIMMERMAN, David (éd.), 1984, *Morality, Reason and Truth*, Totowa: Rowman & Allanheld.
CRAIGHERO, Laila, RIZZOLATTI, Giaccamo, 2004, "The Mirror-Neuron System", *Annual Review of Neuroscience*, 27, pp. 169-192.
CUMMINS, Robert, 1975, "Functional Analysis", *Journal of Philosophy*, 72, pp. 741-764; in E. SOBER (éd.), 1994 (1993), *Conceptual Issues in Evolutionary Biology*, Cambridge (Mass), London: MIT Press, pp. 49-69.
CUSHMAN, Fiery, YOUNG, Liane, HAUSER, Marc, 2006, "The Role of Conscious Reasoning and Intuition on Moral Judgments; Testing Three Principles of Harm", *Psychological Science*, pp. 476-477,
DAMASIO, Antonio, 2001 (1994), *L'erreur de Descartes; la raison des émotions*, trad. de l'angl. par Marcel BLANC, Paris: Odile Jacob.
DANIELS, Norman, 1996, *Justice and Justification; Reflective Equilibrium in Theory and Practice*, Cambridge: Cambridge University Press.
D'AQUIN, Thomas, 1985/1265-1273, *Somme théologique*, Vol. 3 (II-II), Paris: Editions du Cerf.
D'ARMS, Justin, JACOBSON, Daniel, 1994, "Expressivism, Morality, and the Emotions", *Ethics*, 104/4, pp. 739-763.
D'ARMS, Justin, JACOBSON, Daniel, 2000, "Sentiment and Value", *Ethics*, 110/4, pp. 722-748.
DARWIN, Charles, 1859, *L'origine des espèces*, trad. de l'angl. par E. BARBIER. Accessible en ligne: ABU. <http://abu.cnam.fr/> (réf. 10.01.2006)
DARWIN, Charles, 2000 (1871), *La filiation de l'homme et la sélection liée au sexe*, trad. de l'angl. par M. PRUM, Paris: Syllepse.
DAVID, Patrice, SAMADI, Sarah, 2000, *La théorie de l'évolution; Une logique pour la biologie*, Paris: Flammarion.
DAVIS, Jody, RUSBULT, Caryl, 2001, "Attitude Alignment in Close Relationships", *Journal of Personality and Social Psychology*, 81, pp. 65-84.
DAWKINS, Richard, 1979, "Twelve Misunderstandings of Kin Selection", *Zeitschrift für Tierpsychologie*, 51, pp. 184-200.
DAWKINS, Richard, 1996 (1976), *Le Gène égoïste*, trad. de l'angl. par L. OVION, Paris: Odile Jacob.
DAWKINS, Richard, 1999 (1982), *The Extended Phenotype; The long Reach of the Gene*, Oxford: Oxford University Press.
DAWKINS, Richard, 2004, "Extended Phenotype - But Not Too Extended; A Reply to Laland, Turner and Jablonka", *Biology and Philosophy*, 19/3, pp. 377-396.
DEACON, Terrence, 1997, *The symbolic species*, New York: W. W. Norton.
DEHNER, Klaus, 1998, *Lust an Moral; die natürliche Sehnsucht nach Werten*, Darmstadt: Primus Verlag.
DENNETT, Daniel, 2000 (1995), *Darwin est-il dangereux?*, trad. de l'angl. par P. ENGEL, Paris: Odile Jacob.
DEONNA, Julien, 2007, "Evolution, émotion et morale ; Les exemples de la honte et de la culpabilité", in C. CLAVIEN *et al.* (éds.), *Morale et évolution biologique ; entre déterminisme et liberté*, Lausanne: PPUR, pp. 142-164.
DE QUERVAIN, Dominique, FISCHBACHER, Urs, TREYER, Valérie, SCHELLHAMMER, Melanie, SCHNYDER, Ulrich, BUCK, Alfred, FEHR, Ernst, 2004, "The Neural Basis of Altruistic Punishment", *Science*, 305, pp. 1254-1258.
DE SOUSA, Ronald, 1987, *The Rationality of Emotion*, Cambridge (MA): MIT Press.

DE SOUSA, Ronald, 2001, "Moral Emotions", *Ethical Theory and Moral Practice*, 4, pp. 109-126.
DE SOUSA, Ronald, 2003, "Emotion", *Stanford Encyclopedia of Philosophy* (en ligne), pp. 1-23. <http://www.plato.stanford.edu> (réf. 10.12.2006)
DE SOUSA, Ronald, 2004, *Evolution et rationalité*, Paris: PUF.
DE SOUSA, Ronald, 2007, "A propos de quelques défaillances de la Providence", in C. CLAVIEN *et al.* (éds.), *Morale et évolution biologique ; entre déterminisme et liberté*, Lausanne: PPUR, pp. 89-100.
DE VRIES, Hugo, 1900, "Sur la loi de disjonction des Hybrides", *Comptes Rendus de 1'Académie des Sciences*, 130, p. 845-847.
DE WAAL, Frans, 1989, *Peacemaking Among Primates*, Cambridge (Mass): Harward University Press.
DE WAAL, Frans, 1997 (1996), *Le bon singe; les bases naturelles de la morale*, trad. de l'angl., Paris: Bayard.
DE WAAL, Frans, 2000, "Attitudinal Reciprocity in Food Sharing Among Brown Capuchin Monkeys", *Animal Behaviour*, 60, p. 253-261.
DEWEY, John, 1994, *The Moral Writings of John Dewey*, J. GOUINLOCK (éd.), Buffalo, New York: Prometheus Books.
DOBZHANSKY, Theodosius, 1937, *Genetics and the Origin of Species*, New York: Columbia University Press.
DOERING, Sabine, 2007, "Seeing what to Do; Affective Perception and Rational Motivation", *Dialectica*, 61/3, pp. 363-394.
DOSTOÏEVSKI, Fedor, 1996 (1866), *Crime et châtiment*, trad. du russe par P. PASCAL, Paris: GF Flammarion.
DOVER, Kenneth, 1974, *Greek Popular Morality in the Time of Plato and Aristotle*, Oxford: Blackwell.
DUMOUCHEL, Paul, 2004, "Y a-t-il des sentiments moraux ? ", *Dialogue*, 43/3, pp. 471-489.
DUNBAR, Robin, 1996, *Grooming, Gossip, and the Evolution of Language*, Cambridge (Mass): Harvard University Press.
DUNBAR, Robin, 2000, "On the Origin of the Human Mind", in P. CARRUTHERS *et al.* (éds.), 2000, *Evolution and the Human Mind; Modularity, Language and Meta-Cognition*, Cambridge (UK), New York: Cambridge University Press, pp. 238-253.
EDWARDS, Kari, VON HIPPEL, William, 1995, "Hearts and Minds; The Priority of Affective versus Cognitive Factors in Person Perception", *Personality and Social Psychology Bulletin*, 21, pp. 996-1011.
EIBL-EIBESFELDT, Irenäus, 1990 (1988), *Der Mensch – das riskierte Wesen; Zur Naturgeschichte menschlicher Unvernunft*, München, Zürich: Piper.
EISENBERG, Nancy, 1986, *Altruistic Emotion, Cognition, and Behavior*, Hillsdale, New York: Erlbaum.
EISENBERG, Nancy, 2006, "Empathy-Related Responding and Prosocial Behaviour", *Novartis Foundation Symposia*, issue: *Empathy and Fairness*, pp. 71-88.
EKMAN, Paul, FRIESEN, Wallace, 1989, "The Argument and Evidence about Universals in Facial Expressions of Emotion", in H. WAGNER *et al.* (éds.), *Handbook of Social Psychophysiology*, New York: John Wiley and Sons, pp. 143-164.
ELZANOWSKI, Andrzej, 1993, "The Moral Career of Vertebrate Values", in M. NITECKI *et al.* (éds.), *Evolutionary Ethics*, Albany: SUNY Press, pp. 259-276.
ENQUIST, Magnus, LEIMAR, Olof, 1993, "The Evolution of Cooperation in Mobile Organisms", *Animal Behaviour*, 45, pp. 747-757.
FARBER, Paul, 1994, *The Temptation of Evolutionary Ethics*, Berkeley: University of California Press.
FAUCHER, Luc, 2007, "Les émotions morales à la lumière de la psychologie évolutionniste ; le dégoût et l'évitement de l'inceste", in C. CLAVIEN *et al.* (éds.), *Morale et évolution biologique ; entre déterminisme et liberté*, Lausanne: PPUR, pp. 108-141.
FAUCHER, Luc, MACHERY, Edouard, 2004, "Construction sociale, biologie et évolution culturelle ; un modèle intégratif de la pensée raciale", *Archives de l'Institut Jean Nicod*

(en ligne). <http://jeannicod.ccsd.cnrs.fr/ijn_00000532/en/> (réf. 08.06.2006)
FEHR, Ernst, 2004, « Don't Lose Your Reputation », *Nature*, 432, pp. 449-450.
FEHR, Ernst, FISCHBACHER, Urs, 2003, "The Nature of Human Altruism", *Nature*, 425, pp. 785-791.
FEHR, Ernst, FISCHBACHER, Urs, 2004a, "Social Norms and Human Cooperation", *Trends in Cognitive Sciences*, 8/4, pp. 185-190.
FEHR, Ernst, FISCHBACHER, Urs, 2004b, "Third-party Punishment and Social Norms", *Evolution and Human Behavior*, 25, pp. 63-87.
FEHR, Ernst, GAECHTER, Simon, 1998, "Reciprocity and Economics; The Economic Implications of Homo Reciprocans", *European Economic Review*, 42, pp. 845–859.
FEHR, Ernst, GAECHTER, Simon, 2002, "Altruistic Punishment in Humans", *Nature* 415, pp. 137-140.
FEHR, Ernst, GAECHTER, Simon, 2004, "Egalitarian Motive and Altruistic Punishment", *Nature*, pp. E1-E2.
FEHR, Ernst, ROCKENBACH, Bettina, 2003, "Detrimental Effects of Sanctions on Human Altruism", *Nature*, 422, pp. 137-140.
FEINBERG, Joel, 1984, "Psychological Egoism", in S. CAHN *et al.* (éds.), *Reason at Work*, San Diego: Harcourt Brace and Jovanovich, pp. 25-35.
FELDMAN, Fred, 1986, *Doing the Best We Can*, Dordrecht: Reidel.
FESSLER, Daniel, HALEY, Kevin, 2003, "The Strategy of Affect; Emotions in Human Cooperation", *in* P. HAMMERSTEIN (éd.), *Genetic and Cultural Evolution of Cooperation*, Cambridge: MIT Press, pp. 7-36.
FESSLER, Daniel, NAVARRETE, David, 2003, "Meat Is Good to Taboo", *Journal of Cognition and Culture*, 3/1, pp. 1-40.
FISCHBACHER, Urs, GAECHTER, Simon, FEHR, Ernst, 2001, "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment", *Economics Letters*, 71, pp. 397-404.
FISHER, James, HINDE, Robert, 1949, "Further Observations on the Opening of Milk Bottles by Birds", *British Birds*, 42, pp. 347-357.
FISHER, Ronald, 1930, *The Genetical Theory of Natural Selection*, Oxford: Clarendon Press.
FLACK, Jessica, DE WALL, Frans, 2000, "Any Animal Whatever; Darwinian Building Blocks of Morality in Monkeys and Apes", *Journal of Consciousness Studies*, 7, L. KATZ, ed., *Evolutionary Origins of Morality*, Bowling Green: Imprint Academic, pp. 1-29.
FLINN, Mark, ALEXANDER, Richard, 1982, "Culture theory; The Developing Synthesis from Biology", *Human Ecology*, 10/3, pp. 383-400.
FOOT, Philippa, 1967, "The Problem of Abortion and the Doctrine of Double-Effect", *Oxford Review*, 5, pp. 5-15.
FOOT, Philippa, 2001, *Natural Goodness*, Oxford, New York: Oxford University Press.
FOOT, Philippa, 2002, *Moral Dilemmas and other Topics in Moral Philosophy*, Oxford: Clarendon Press.
FOWLER, James, 2004, "Altruistic Punishment and the Origin of Cooperation", *Ideas Repec* (en ligne). <http://ideas.repec.org/p/wpa/wuwpga/0410002.html> (réf. 16.05.2007)
FOWLER, James, JOHNSON, Tim, SMIRNOV, Oleg, 2004, "Egalitarian Motive and Altruistic Punishment", *Nature*, pp. E1-E2.
FOWLER, James, JOHNSON, Tim, MCELREATH, Richard, SMIRNOV, Oleg, 2005, "Egalitarian Punishment in Humans", *Ideas Repec* (en ligne). <http://ideas.repec.org/p/wpa/wuwpex/0507003.html> (réf. 16.05.2007)
FOX, Robin (éd.), 1975, *Biosocial Anthropology*, London: Malaby Press.
FRANK, Robert, 1988, *Passions Within Reason*, New York, London: W. W. Norton & Company.
FRANKENA, William, 1988 (1967), "The Naturalistic Fallacy", in P. FOOT (éd.), 1988 (1967), *Theories of Ethics*, Oxford: Oxford University Press, pp. 50-63.
FREDRICKSON, Barbara, 2000, "Cultivating Positive Emotions to Optimize Health and Well-Being", *Prevention & Treatment*, 3, pp. 1-25.

FREDRICKSON, Barbara, 2003, "The Value of Positive Emotions", *American Scientist*, 91, pp. 330-335.
GAECHTER, Simon, FALK, Armin, 2002, "Reputation and Reciprocity; Consequences for the Labour Relation", *Scandinavian Journal of Economics*, 104/1, pp. 1-27.
GALLESE, Vittorio, FADIGA, Luciano, FOGASSI Leonardo, RIZZOLATTI, Giaccamo, 1996, "Action Recognition in the Premotor Cortex", *Brain*, 119, pp. 593-609.
GALLESE, Vittorio, GOLDMAN, Alvin, 1998, "Mirror Neurons and the Simulation Theory of Mind-Reading", *Trends of Cognitive Sciences*, 12, pp. 493–501.
GALLESE, Vittorio, KEYSERS, Christian, RIZZOLATTI, Giaccamo, 2004, "A Unifying View of the Basis of Social Cognition", *Trends in Cognitive Sciences*, 8/9, pp. 396-403.
GALLUP, Gordon, 1977, "Self-Recognition in Primates; A Comparative Approach to the Bidirectional Properties of Consciousness", *American Psychologist*, 32, pp. 329-338.
GARDNER, Andy, WEST, Stuart, 2004, "Cooperation and Punishment, Especially in Humans", *The American Naturalist*, 164/6, pp. 753-764.
GAYON, Jean, 1998, *Darwinism's Struggle for Survival; Heredity and the Hypothesis of Natural Selection*, Cambridge: Cambridge University Press, 1998.
GAYON, Jean, 1999, "Sélection naturelle biologique et selection naturelle économique ; Examen philosophique d'une analogie", *Économies et Sociétés*, Hors Série, 35/1, pp. 107-126.
GAYON, Jean, 2000, "From Measurement to Organization; a Philosophical Scheme for the History of the Concept of Heredity", in P. BEURTON *et al.* (éds.), *The Concept of the Gene in Development and Evolution; Historical and Epistemological Perspectives*, Cambridge: Cambridge University Press, pp. 69-90
GAYON, Jean, 2002, "Y a-t-il un concept biologique de la race ?", *Annales d'histoire et de philosophie du vivant*, 6, pp. 155-176.
GEACH, Peter, 1965, "Assertion", *Philosophical Review*, 74, pp. 449–65.
GEIGER, Gebhard, 1992, "Why There Are No Objective Values; A Critique of Ethical Intuitionism From Evolutionary Point of View", *Biology and Philosophy*, 7, pp. 315-330.
GIBBARD, Allan, 2002 (1990), *Wise Choices, Apt Feelings; A Theory of Normative Judgment*, Oxford: Oxford University Press.
GILDENHUYS, Peter, 2003, "The Evolution of Altruism; The Sober/Wilson Model", *Philosophy of Science*, 70, pp. 27-48.
GINTIS, Herbert, 2002 (2000), "Group Selection and Human Prosociality", in L. KATZ (éd.), 2002 (2000), *Evolutionary Origins of Morality; Cross-Disciplinary Perspectives*, Bowling Green: Imprint Academic, pp. 215-219.
GINTIS, Herbert, 2000, "Strong reciprocity and human sociality", J*ournal of Theoretical Biology*, 206, pp. 169–179.
GINTIS, Herbert, 2003, "The Hitchhiker's Guide to Altruism; Gene-Culture Coevolution, and the Internalization of Norms", *Journal of Theoretical Biology*, 220, pp. 407-418.
GINTIS, Herbert, BOWLES, Samuel, BOYD, Robert, FEHR, Ernst, 2003, "Explaining Altruistic Behavior in Humans", *Evolution and Human Behavior*, 24, pp. 153-172.
GINTIS, Herbert, SMITH, Eric, BOWLES, Samuel, 2001, "Costly Signalling and Cooperation", *Journal of Theoretical Biology*, 213, pp. 103-119.
GOLDIE, Peter, 2000, *The Emotions; A Philosophical Exploration*, Oxford: Clarendon Press.
GOLDIE, Peter, 2007, "Seeing what is the Kind Thing to Do; Perception and Emotion in Morality", *Dialectica*, 61, pp. 347–361.
GOLDMAN, Alan, 1990, *Moral Knowledge*, London, New York: Routledge.
GOULD, Stephen, 1999, *Rocks of Ages; Science and Religion in the Fullness of Life*, New York: Ballantine Books.
GOULD, Stephen, LEWONTIN, Richard, 1979, "The Spandrels of San Marco and the Panglossian Paradigm; a Critique of the Adaptationist Programme", *Proceedings of the Royal Society of Londondon B*, 205, pp. 581-598.
GRAFEN, Alan, 1985, "A Geometric View of Relatedness", in R. DAWKINS *et al.* (éds.), *Oxford Surveys in Evolutionary Biology*. Vol. 2, Oxford: Oxford University Press, pp. 28-89.

GREENE, Joshua, 2005, "Cognitive Neuroscience and the Structure of the Moral Mind", in P. CARRUTHERS *et al.* (éds), *The Innate Mind; Structure and Content*, Oxford: Oxford University Press, pp. 338-352.
GREENE, Joshua, HAIDT, Jonathan, 2002, "How (and Where) does Moral Judgment Work?", *Trends in Cognitive Sciences*, 6, pp. 517-523.
GREENE, Joshua, SOMMERVILLE, Brian, NYSTROM, Leigh, DARLEY, John, COHEN, Jonathan, 2001, "An fMRI Investigation of Emotional Engagement in Moral Judgment", *Science*, 293, pp. 2105-2108.
GRIFFIN, Ashleigh, WEST, Stuart, BUCKLING, Angus, 2004, "Cooperation and Competition in Pathogenic Bacteria", *Nature*, 430, pp. 1024-1027.
HAIDT, Jonathan, 2000, "The Positive Emotion of Elevation", *Prevention & Treatment*, 3, pp.1-4.
HAIDT, Jonathan, 2001, "The Emotional Dog and Its Rational Tail; A Social Intuitionist Approach to Moral Judgment", *Psychological Review*, 108/4, pp. 814-834.
HAIDT, Jonathan, 2003, "The Moral Emotions", in R. DAVIDSON *et al.* (éds.), *Handbook of Affective Sciences*, pp. 852-870.
HAIDT, Jonathan, JOSEPH, Craig, 2004, "Intuitive Ethics; How Innately Prepared Intuitions Generate Culturally Variable Virtues", *Deadalus, Special Issue on Human Nature*, pp. 55-66.
HALDANE, John, 1932, *The Causes of Evolution*, London: Longman.
HALEY, Kevin, FESSLER, Daniel, 2005, "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game", *Evolution and Human Behavior*, 26, pp. 245-256.
HAMILTON, William, 1963, "The Evolution of Altruistic Behavior", *The American Naturalist*, 97/892, pp. 354-356.
HAMILTON, William, 1964, "The Genetical Evolution of Social Behaviour (I and II)", *Journal of Theoretical Biology*, 7, pp. 1-52.
HAMILTON, William, 1970, "Selfish and Spiteful Behaviour in an Evolutionary Model", *Nature*, 228, pp. 1218-1220.
HAMILTON, William, 1975, "Innate Social Aptitudes of Man; an Approach from Evolutionary Genetics", in R. FOX (éd.), *Biosocial Anthropology*, London: Malaby Press, pp. 133-155.
HAMMERSTEIN, Peter (éd.), 2003a, *Genetic and Cultural Evolution of Cooperation*, Cambridge: MIT Press.
HAMMERSTEIN, Peter, 2003b, "Why Is Reciprocity So Rare in Social Animals?", in P. HAMMERSTEIN (éd.), *Genetic and Cultural Evolution of Cooperation*, Cambridge: MIT Press, pp. 83-94.
HARDIN, Garrett, 1968, "The Tragedy of the Commons", *Science*, 162, pp. 1243-1248.
HARMAN, Gilbert, 1977, *The Nature of Morality*, New York: Oxford University Press.
HARMAN, Gilbert, 1999, "Moral Philosophy and Linguistics", in K. BRINKMANN (éd.), *Proceedings of the 20th World Congress of Philosophy*, Volume I: *Ethics*. Bowling Green: Philosophy Documentation Center, pp. 107-115; in G. HARMAN, 2000, *Explaining Value; And Other Essays in Moral Philosophy*, pp. 217-226.
HARMAN, 2000, *Explaining Value; And Other Essays in Moral Philosophy*, Oxford: Oxford University Press.
HARMS, William, 2000, "Adaptation and Moral Realism", *Biology and Philosophy*, 15, pp. 699-712.
HATFIELD, Elaine, CACIOPPO, John, RAPSON, Richard, 1994, *Emotional Contagion*, New York: Cambridge University Press.
HAUBER, Mark, SHERMAN, Paul, 1998, "Nepotism and Marmot Alarm Calling", *Animal Behaviour*, 56, pp. 1049-1052.
HAUSER, Marc, 2006, *Moral Minds*, New York: Harper Collins.
HELWIG, Charles, TURIEL, Elliot, 2002, "Children's Social and Moral Reasoning", in P. SMITH *et al.* (éds.), *Blackwell Handbook of Childhood Social Development*, Oxford: Blackwell, pp. 475-490.
HENRICH, Joseph, BOYD, Robert, 1998, "The Evolution of Conformist Transmission and the Emergence of Between-Group Differences", *Evolution and Human Behavior*, 19, pp. 215-

241.
HENRICH, Joseph, BOYD, Robert, 2001, "Why People Punish Defectors; Weak Conformist Transmission can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas", *Journal of Theoretical Biology*, 208, pp. 79-89.
HENRICH, Joseph, BOYD, Robert, 2002, "On Modeling Cognition and Culture; Why Replicators are not Necessary for Cultural Evolution », *Journal of Cognition and Culture*, 2, pp. 87-112.
HENRICH, Joseph, BOYD, Robert, BOWLES, Samuel, CAMERER, Colin, (éds.), 2004, *Foundations of Human Sociality; Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, Oxford: Oxford University Press
HENRICH, Joseph, GIL-WHITE, Francisco, 2001, "The Evolution of Prestige Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission", *Evolution and Human Behavior*, 22, pp. 165-196.
HENRICH, Joseph, MCELREATH, Richard, 2003, "The Evolution of Cultural Evolution", *Evolutionary Anthropology*, 12, pp. 123-135.
HOBBES, Thomas, 2000 (1651), *Léviathan*, Paris: Gallimard.
HULL, David, 1980, "Individuality and Selection", *Annual Review of Ecology and Systematics*, 11, pp. 311-332.
HUME, David, 1946 (1740), *Traité de la Nature Humaine*, T.2, Paris: Aubier.
HUNEMAN, Philippe, 2007, "L'éthique évolutionniste version déflationniste, ou la sagesse des Dowayos", in C. CLAVIEN *et al.* (éds.), *Morale et évolution biologique ; entre déterminisme et liberté*, Lausanne: PPUR, pp. 245-265.
HUTCHESON, Francis, 1991 (1726), *Recherche sur l'origine de nos idées de la beauté et de la vertu*, trad. de l'angl. par A.-D. BALMES, Paris: Vrin.
HUXLEY, Thomas, 1893, "Evolution and Ethics", in 1894, *Collected Essays*, Vol. IX, London: Macmillan, pp. 46-116 ; in H. NITECKI *et al.* (éds.), 1993, *Evolutionary Ethics*, Albany: SUNY Press, pp. 29-80. Accessible en ligne: <http://aleph0.clarku.edu/huxley/CE9/E-E.html> (réf. 19.06.2007)
IRONS, William, 1979, "Cultural and Biological Success", in W. IRONS *et al.* (éds.), *Evolutionary Biology and Human Social Behavior*, North Scituate: Duxbury Press, pp. 257-272
JACKSON, Frank, 1998, *From Metaphysics to Ethics; a Defence of Conceptual Analysis*, Oxford: Clarendon Press.
JACKSON, Frank, PETTIT, Philip, 1995, "Moral Functionalism and Moral Motivation", *The Philosophical Quarterly*, 45/178, pp. 20-40.
JACKSON, Frank, PETTIT, Philip, 1996, "Moral Functionalism, Supervenience and Reductionism", *The Philosophical Quarterly*, 46/182, pp. 82-86.
JACOB, Pierre, JEANNEROD, Marc, 2005, "The Motor Theory of Social Cognition; a Critique", *Trends in Cognitive Sciences*, 9/1, pp. 21-25.
JAMIESON, Dale, 2002, "Sober and Wilson on Psycholgical Altruism", *Philosophy and Phenomenological Research*, 65/3, pp. 702-710.
JOHANNSEN, Wilhelm, 1909, *Elemente der exakten Erblichkeitslehre*, Jena: Gustav Fischer.
JOHANNSEN, Wilhelm, 1911, "The Genotype Conception of Heredity", *American Naturalist*, 45, pp. 129-159.
JONAS, Hans, 2001/1990, *Le principe responsabilité*, trad. de l'all. Par J. GREISCH, Paris: Flammarion.
JONSEN, Albert, TOULMIN, Stephen, 1988, *The Abuse of Casuistry; A History of Moral Reasoning*. Berkeley: University of California Press.
JOYCE, Richard, 2000, "Darwinian Ethics and Error", *Biology and Philosophy*, 15, pp. 713-732.
JOYCE, Richard, 2006, *The Evolution of Morality*, Cambridge: MIT Press.
KANT, Emmanuel, 1997 (1785), *Fondements de la métaphysique des moeurs*, trad. de l'all. par V. DELBOS revue par A. PHILONENKO, Paris: Vrin.
KANT, Emmanuel, 1997 (1788), *Critique de la raison pratique*, trad. de l'all. par F. PICAVET, Paris: Quadrige / PUF.

KATZ, Leonard (éd.), 2002 (2000), *Evolutionary Origins of Morality; Cross-Disciplinary Perspectives*, in *Journal of Consciousness Studies*, 7, Bowling Green: Imprint Academic.
KELLER, Laurent, ROSS, Kenneth, 1998, "Selfish Genes; A Green Beard in the Red Fire Ant", *Nature*, 394, pp. 573-575.
KELLY, Daniel, STICH, Stephen, 2008, "Two Theories About the Cognitive Architecture Underlying Morality", in P. CARRUTHERS *et al.* (éds.), *The Innate Mind; Foundations and the Future*, pp.
KELLY, Daniel, STICH, Stephen, HALEY, Kevin, ENG, Serena, FESSLER, Daniel, 2007, "Harm, Affect, and the Moral / Conventional Distinction", *Mind & Language*, 22/2, pp. 117–131.
KITCHER, Philip, 1987 (1985), *Vaulting Ambition; Sociobiology and the Quest for Human Nature*, Cambridge (US), London: MIT Press.
KITCHER, Philip, 1994 (1993), "Four Ways of 'biologicizing' Ethics", in E. SOBER (éd.), 1994 (1993), *Conceptual Issues in Evolutionary Biology*, Cambridge (US), London: MIT Press, pp. 439-450; in P. KITCHER, 2003, *In Mendel's Mirror*, New York: Oxford University Press, pp. 321-332.
KITCHER, Philip, 1998, "Psychological Altruism, Evolutionary Origins, and Moral Rules", *Philosophical Studies*, 89, pp. 283-316.
KITCHER, Philip, 2006, "Between Fragile Altruism and Morality; Evolution and the Emergence of Normative Guidance", in G. BONIOLO *et al.* (éds.), *Evolutionary Ethics and Contemporary Biology*, Cambridge, New York: Cambridge University Press, pp. 159-177.
KNIGHT, Nicola, à paraître, "The Action Classification Approach to the Psychology of Normative Violations".
KREBS, Dennis, 2002 (2000), "As Moral as We Need to Be", in L. KATZ (éd.), 2002 (2000), *Evolutionary Origins of Morality; Cross-Disciplinary Perspectives*, in *Journal of Consciousness Studies*, 7, Bowling Green: Imprint Academic, pp. 139-143.
KREBS, Dennis, RUSSELL, Cristine, 1981, "Role-Taking and Altruism; When you put yourself in the Shoes of Another, will they Take You to their Owner's Aid? " in J. RUSHTON *et al.* (éds.), *Altruism and Helping Behavior; Social, Personality, and Developmental Perspectives*, Hillsdale, New-York: Lawrence Erlbaum Associates, pp. 137-165.
KREBS, John, DAVIES, Nicholas, 1993 (1981), *An Introduction to Behavioural Ecology*, Oxford, London: Blackwell.
KRISTJANSSON, Kristjan, 2002, *Justifying Emotions; Pride and Jealousy*, London, New York: Routledge.
KUPIEC, Jean-Jacques, SONIGO, Pierre, 2000, *Ni Dieu ni gène*, Paris: Seuil.
LACHAPELLE, Jean, FAUCHER, Luc, POIRIER, Pierre, 2006, "Cultural Evolution, the Baldwin Effect, and Social Norms", in N. GONTIER *et al.* (éds.), *Evolutionary Epistemology, Language, and Culture* (en ligne), Brussels: Vrije Universiteit Brussel, pp. 313-334.
LAHTI, David, 2003, "Parting with Illusions in Evolutionary Ethics", *Biology and Philosophy*, 18, pp. 639-651.
LALAND, Kevin, 2004, "Extending the Extended Phenotype", *Biology and Philosophy*, 19/3, pp. 313-325.
LANGANEY, André, 2001, *La philosophie biologique*, Paris: Editions Belin.
LEHMANN, Laurent, KELLER, Laurent, 2006, "The Evolution of Cooperation and Altruism - A General Framework and a Classification of Models", *Journal of Evolutionary Biology*, 19, pp. 1365-1376.
LEHMANN, Laurent, PERRIN, Nicolas, 2002, "Altruism, Dispersal, and Phenotype-Matching Kin Recognition", *The American Naturalist*, 159, pp. 451-468.
LEHMANN, Laurent, PERRIN, Nicolas, ROUSSET, François, 2006, "Population Demography and the Evolution of Helping Behaviors", *Evolution*, 60/6, pp. 1137-1151.
LEIMAR, Olof, CONNOR, Richard, 2003, "By-product Benefits, Reciprocity, and Pseudoreciprocity in Mutualism", in P. HAMMERSTEIN (éd.), *Genetic and Cultural*

Evolution of Cooperation, Cambridge: MIT Press, pp. 203-222.
LEIMAR, Olof, HAMMERSTEIN, Peter, 2001, "Evolution of Cooperation through Indirect Reciprocity", *Proceedings of the Royal Society of London B*, 268, pp. 745-753.
LEMOS, John, 1999, "Bridging the Is/Ought Gap with Evolutionary Biology; Is This a Bridge Too Far?", *The Southern Journal of Philosophy*, 37/4, pp. 559-577.
LESTEL, Dominique, 2003 (2001), *Les origines animales de la culture*, Paris: Flammarion.
LEWIS, Michael, BROOKS-GUNN, Jeanne, 1981, "Le développement de la reconnaissance de soi", *in* P. MOUNOUD *et al.* (éds.), *La reconnaissance de son image chez l'enfant et l'animal*, Neuchâtel: Delachaux et Niestlé.
LEWONTIN, Richard, 1970, "The Units of Selection", *Annual Review of Ecology and Systematics*, 1, pp. 1-18.
LORENZ, Konrad, 1977 (1963), *L'agression ; une histoire naturelle du mal*, trad. de l'all. par V. FRITSCH, Paris: Flammarion.
LORENZ, Konrad, 1989 (1988), *Les oies cendrées*, trad. de l'all. par C. DHORBAIS, Paris: Albin Michel.
LUHMANN, Niklas, 1987, *Soziale Systeme; Grundriss einer allgemeinen Theorie*, Frankfurt am Main: Suhrkamp.
LUMSDEN, Charles, WILSON, Edward, 1983, *Promethean Fire*, Cambridge: Harvard University Press
MACHERY, Edouard, 2003, "Culture et singularité humaine", *Archives de l'Institut Jean Nicod* (en ligne). <http://jeannicod.ccsd.cnrs.fr> (réf. 25.05.2005)
MACKIE, John, 1977, *Ethics; Inventing Right and Wrong*, New York: Penguin Books.
MACKIE, John, 1989, "The Law of the Jungle; Moral Alternatives and Principles of Evolution", in M. RUSE (éd.), *Philosophy of Biology*, New York, London: Macmillan, pp. 303-312.
MALECOT, Gustave, 1948, *Les mathématiques de l'hérédité*, Paris: Masson.
MAMELI, Mateo, 2005 (2004), "The Role of Emotions in Ecological and Practical Rationality", in D. EVANS *et al.* (eds.), 2005 (2004), *Emotion, Evolution and Rationality*, Oxford: Oxford University Press, pp. 159-178.
MANDEVILLE, Bernard, 1990 (1714), *La fable des abeilles*, Paris: Vrin.
MANSON, Joseph, NAVARRETE, David, SILK, Joan, PERRY, Susan, 2004, "Time-Matched Grooming in Female Primates?", *Animal Behaviour*, 67, pp. 493-500.
MARWELL, Gerald, AMES, Ruth, 1981, "Economists Free Ride; Does Anyone Else?", *Journal of Public Economics*, 15, pp. 295-310.
MASLOW, Abraham, 1943, "A Theory of Human Motivation", *Psychological Review*, 50, pp. 370-396.
MASSERMAN, Jules, WECHKIN, Stanley, TERRIS, William, 1964, "'Altruistic' Behavior in Rhesus Monkeys", *American Journal of Psychiatry*, 121, pp. 584-585.
MATEO, Jill, 2003, "Kin Recognition in Ground Squirrels and Other Rodents", *Journal of Mammalogy*, 84/4, pp. 1163-1181.
MATZKE, Marjori, MATZKE, Antonius, KOOTER, Jan, 2001, "RNS; Guiding Gene Silencing", *Science*, 293, pp. 1080-1083.
MAYNARD SMITH, John, 1964, "Group Selection and Kin Selection", *Nature*, 201, pp. 1145-1147.
MAYNARD SMITH, John, 1976, "Group Selection", *The Quarterly Review of Biology*, 51/2, pp. 277-283.
MAYNARD SMITH, John, 1982, *Evolution and the Theory of Games*, Cambridge: Cambridge University Press.
MAYNARD SMITH, John, 1993, *The Theory of Evolution*, Cambridge: Cambridge University Press.
MAYNARD SMITH, John, 1998, "The Origin of Altruism", *Nature*, 393, pp. 639-340.
MAYNARD SMITH, John, 2001 (1998), *La construction du vivant; gènes, embryons et évolution*, trad. de l'angl. par A. LESNE, Paris: Cassini (coll. Le sel et le fer).
MAYNARD SMITH, John, HARPER, David, 2003, *Animal Signals*, Oxford: Oxford University Press.
MAYNARD SMITH, John, PRICE, George, 1973, "The Logic of Animal Conflicts", *Nature*,

246, pp. 15-18.
MAYNARD SMITH, John, SZATHMARY, Eörs, 2000 (1999), *Les origines de la vie ; de la naissance de la vie à l'origine du langage*, trad. de l'angl. par N. CHEVASSUS-AU-LOUIS, Paris: Dunod.
MAYR, Ernst, 1957, "Species Concepts and Definitions", in E. MAYR (éd.), *The Species Problem, Bulletin of American Society for the Advancement of Science*, 50, pp. 1-22.
MAYR, Ernst, 1961, "Cause and Effect in Biology", *Science*, 134, pp. 1501-1506.
MAYR, Ernst, 1963, *Animal Species and Evolution*, Cambridge: Harvard University Press.
MAYR, Ernst, 1989 (1982), *Histoire de la biologie ; diversité, évolution et hérédité*. trad. de l'angl. par M. BLANC, Paris: A. Fayard.
MCBREARTY, Sally, BROOKS, Alison, 2000, "The Revolution that Wasn't; A New Interpretation of the Origin of Modern Human Behavior", *Journal of Human Evolution*, 39, pp. 453-563.
MCCULLOUGH, Michael, KILPATRICK, Shelley, EMMONS, Robert, LARSON, David, 2001, "Is Gratitude a Moral Affect?", *Psychological Bulletin*, 127/2, pp. 249-266.
MCDOWELL, John, 1985, "Value and Secondary Qualities", in T. HONDERICH (éd.), *Morality and Objectivity; A Tribute to J.L. Mackie*, London: Routledge & Kegan, pp. 110-129; trad. française in R. OGIEN (éd.), 1999, *Le Réalisme Moral*, Paris: PUF.
MCELREATH, Richard, BOYD, Robert, RICHERSON, Peter, 2003, "Shared Norms can Lead to The Evolution of Ethnic Markers", *Current Anthropology*, 44, pp. 122-130.
MCELREATH, Richard *et al.*, 2003, "Group Report; The Role of Cognition and Emotion in Cooperation", in P. HAMMERSTEIN (éd.), *Genetic and Cultural Evolution of Cooperation*, Cambridge: MIT Press, pp. 125-152.
MCSHEA, Robert, 1978, "Human Nature Theory and Political Philosophy", *American Journal of Political Science*, 22, pp. 656-679.
MCSHEA, Robert, MCSHEA, Daniel, 1999, "Biology and Value Theory", in J. MAIENSCHEIN *et al.* (éds.), *Biology and the Foundation of Ethics*, Cambridge: Cambridge University Press, pp. 307-327.
MENDEL, Gregor, 1911 (1865), "Versuche über Pflanzenhybriden", *Verhandlungen des naturforschenden Vereins*, Vol. 4, Brünn: Verlage des Vereins, pp. 3-47. Accessible en ligne: <http://www.biologie.uni-hamburg.de/b-online/d08_mend/mendel.htm> (réf. 05.06.2007)
MIKHAIL, John, 2002, "Aspects of the Theory of Moral Cognition; Investigating Intuitive Knowledge of the Prohibition of Intentional Battery and the Principle of Double Effect", *Social Science Research Network* (en ligne), pp. 1-129. <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=762385> (réf. 21.05.2006)
MILINSKI, Manfred, SEMMANN, Dirk, BAKKER, Theo, KRAMBECK, Hans-Jürgen, 2001, "Cooperation through Indirect Reciprocity; Image Scoring or Standing Strategy?", *Proceedings of the Royal Society of London B*, 268, pp. 2495-2501.
MILINSKI, Manfred, SEMMANN, Dirk, KRAMBECK, Hans-Jürgen, 2002, "Reputation Helps Solve the Tragedy of the Commons", *Nature*, 415, pp. 424-426.
MILL, John Stuart, 1998 (1861), *L'utilitarisme*, trad. de l'angl. par C. AUDARD, Paris: Quadrige / PUF.
MOHR, Hans, 1987, "Evolutionäre Erkenntnistheorie, Ethik und Moral", in R. RIEDL *et al.* (éds.), *Die Evolutionäre Erkenntnistheorie*, Berlin, Hamburg: Verlag Paul Parey, pp. 238-347.
MOLL, Henrike, TOMASELLO, Michael, 2007, "Cooperation and Human Cognition; The Vygotskian Intelligence Hypothesis", *Philosophical Transactions of The Royal Society B; Biological Sciences*, 362, pp. 639-648.
MOLL, Jorge, DE OLIVEIRA-SOUZA, Ricardo, ESLINGER, TOVAR MOLL, Fernanda, AZEVEDO IGNACIO, Fatima, BRAMATI, Ivanei, CAPARELLI-DAQUER, Egas, ESLINGER, Paul, 2005, "The Moral Affiliations of Disgust; a Functional MRI Study", *The Cognitive Behavioral Neurology*, 18/1, pp. 68-78.
MOORE, George, 1903, *Principia Ethica*, Cambridge: Cambridge University Press. Accessible en ligne: <http://fair-use.org/g-e-moore/principia-ethica> (réf. 21.04.2006)

MOSKOWITZ, Gordon, SKURNIK, Ian, GALINSKY, Adam, 1999, "The History of Dual Process Notions, and the Future of Pre-Conscious Control", in S. CHAIKEN *et al.* (éds.), *Dual Process Theories in Social Psychology*, New York: Guilford Press, pp. 12-36.

MURPHY, Jeffrie, 1982, *Evolution, Morality, and the Meaning of Life*, Totowa: Rowman & Littlefield.

NAGEL, Thomas, 1970, *The Possibility of Altruism*, Oxford: Oxford University Press.

NAGEL, Thomas, 1983 (1978), "Ethique sans biologie", in T. NAGEL, *Questions mortelles*, trad. de l'angl. par P. ENGEL, Paris: PUF, pp. 167-172.

NICHOLS, Shaun, 2004, *Sentimental Rules; On the Natural Foundations of Moral Judgment*, Oxford: Oxford University Press.

NIETZSCHE, Friedrich, 2000 (1887), *La généalogie de la morale*, trad. de l'all. par P. WOTLING, Paris: Le Livre de Poche (Classiques de la philosophie).

NISBETT, Richard, COHEN, Dov, 1996, *Culture of Honor; The Psychology of Violence in the South*, Boulder: Westview Press.

NISBETT, Richard, WILSON, Timothy, 1977, "Telling More Than We Can Know; Verbal Reports on Mental Processes", *Psychological Review*, 84, pp. 231–259.

NOË, Ronald, HAMMERSTEIN, Peter, 1994, "Biological Markets: Supply and Demand Determine the Effect of Partner Choice in Cooperation, Mutualism and Mating", *Behavioral Ecology and Sociobiology*, 35, pp. 1-11.

NOWAK, Martin, SIGMUND, Karl, 1998, "Evolution of Indirect Reciprocity by Image Scoring", *Nature*, 393, pp. 573-577.

O'HEAR, Anthony, 1997, *Beyond Evolution; Human Nature and the Limits of Evolutionary Explanation*, Clarendon Press: Oxford.

OKASHA, Samir, 2002, "Genetic Relatedness and the Evolution of Altruism", *Philosophy of Science*, 69, pp. 138-149.

OKLESHEN, Marilyn, 1996, "A Cross Cultural Comparison of Ethical Perspectives and Decision Approaches of Business Students; United States of America versus New Zealand", *Journal of Business Ethics*, 15, pp. 537-549.

OSTROM, Elinor, 1998, "A Behavioral Approach to the Rational Choice Theory of Collective Action; Presidential Address, American Political Science Association, 1997", *American Political Science Review*, 92/1, pp. 1-22.

OSTROM, Elinor, 1990, *Governing the Commons*, New York: Cambridge University Press.

PANCHANATHAN, Karthik, BOYD, Robert, 2004, « Indirect Reciprocity Can Stabilize Cooperation Without the Second-Order Free Rider Problem », *Nature*, 432, pp. 499-502.

PENNISI, Elizabeth, 2001, "Behind the Scenes of Gene Expression", *Science*, 293, pp. 1064-1067.

PEPPER, John, 2000, "Relatedness in Trait Group Models of Social Evolution", *Journal of Theoretical Biology*, 206/3, pp. 355-368.

PERNER, Josef, WIMMER, Heinz, 1985, "'John Thinks That Mary Thinks That...'; Attribution of Second-Order Beliefs by Five- to 10-Year-Old Children", *Journal of Experimental Child Psychology*, 39/3, pp. 437-471.

PERRIN, Nicolas, 2005, *Biologie des populations*, Cours dispensé au semestre d'hiver 2005 à l'Université de Lausanne, faculté de Biologie et Médecine, département d'écologie et d'évolution.

PILIAVIN, Jane, CHARNG, Hong-Wen, 1990, "Altruism; A Review of Recent Theory and Research", *Annual Review of Sociology*, 16, pp. 27-65.

PLATON, 1966 (5-4ème s. av. J.-C.), *La république*, trad. du grec ancien par R. BACCOU, Paris: GF Flammarion. Accessible en ligne: <http://plato-dialogues.org/fr/tetra_4/republic/extraits.htm#call3> (réf. 01.06.2006).

PLATON, 1969 (5-4ème s. av. J.-C.), *Sophiste, Politique, Philèbe ; Timée, Critias*, trad. du grec ancien par E. CHAMBRY, Paris : Garnier Flammarion.

POVINELLI, Daniel, NELSON, Kurt, BOYSEN, Sarah, 1992, "Comprehension of Role Reversal in Chimpanzees; Evidence of Empathy", *Animal Behaviour*, 43, pp. 633–640.

PRESTON, Stephanie, DE WAAL, Frans, 2002 (2001), "Empathy; Its Ultimate and Proximate Bases", *Behavioral and Brain Sciences*, 25/1, pp. 1-72.

PRICE, George, 1970, "Selection and covariance", *Nature*, 227, pp. 529-531.
PRICE, Michael, COSMIDES, Leda, TOOBY, John, 2002, "Punitive Sentiment as an Anti-Free Rider Psychological Device", *Evolution and Human Behavior*, 23, pp. 203-231.
PRINZ, Jesse, 2006, "The Emotional Basis of Moral Judgments", *Philosophical Exploration*, 9/1, pp. 29-43.
PRINZ, Jesse, 2007, *The Emotional Construction of Morals*, Oxford: Oxford University Press.
PRINZ, Jesse, 2008, "Against Moral Nativism", in D. MURPHY *et al.* (éds.), *Stich and his Critics*, Oxford: Blackwell, pp.
PROUST, Joëlle, 1995, "Fonction et causalité", *Intellectica*, 21, pp. 81-113.
PROUST, Joëlle, 2003, *Les animaux pensent-ils?*, Paris: Bayard (Le temps d'une question).
PUTNAM, Hilary, 1981, *Reason, Truth and History*, Cambridge: Cambridge University Press.
PUTNAM, Hilary, 2004 (2002), *Fait/Valeur ; la fin d'un dogme*, trad. de l'angl. par J.-P. CAVERIBÈRE, Paris, Tel-Aviv: Editions de l'Eclat.
PUTNAM, Hilary, 2004, *Ethics Without Ontology*, Cambridge (Mass): Harvard University Press.
QUELLER, David, 1992, "Does Population Viscosity Promote Kin Selection? " *Trends in Ecology and Evolution*, 7, pp. 322-4.
QUELLER, David, 1994, "Genetic Relatedness in Viscous Populations", *Evolutionary Ecology*, 8, pp. 70-73.
QUELLER, David, 2001, "W. D. Hamilton and the Evolution of Sociality", *Behavioral Ecology*, 12/3, pp. 261-268.
QUINE, Willard, 1986, "Reply to Morton White", in L. HAHM *et al*. (éds.), *The Philosophy of WVO Quine*, La Salle: Open Court, pp. 663-665.
RACHELS, James, 1990, *Created from Animals; The Moral Implications of Darwinism*, Oxford, New York: Oxford University Press.
RAUSCHER, Frederick, 1997, "How a Kantian Can Accept Evolutionary Metaethics", *Biology and Philosophy*, 12, pp. 303-326.
RAWLS, John, 1997 (1971), *Théorie de la Justice*, trad. de l'angl. par C. AUDARD, Paris: Seuil.
RICHARDS, Robert, 1986, "A Defense of Evolutionary Ethics", *Biology and Philosophy*, 1, pp. 265-293; in R. RICHARDS, 1987, *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*, pp. 595-627.
RICHARDS, Robert, 1993, "Birth, Death, and Resurrection of Evolutionary Ethics", in H. NITECKI *et al.* (éds.), *Evolutionary Ethics*, Albany: SUNY Press, pp. 113-131.
RICHARDS, Robert, 1999, "Darwin's Romantic Biology; The Foundation of His Evolutionary Ethics", in J. MAIENSCHEIN *et al.* (éds.), *Biology and the Foundation of Ethics*, Cambridge: Cambridge University Press, pp. 113-153.
RICHERSON, Peter, BOYD, Robert, 2000, "Climate, Culture and the Evolution of Cognition", in C. HEYES *et al.* (éds.), *The Evolution of Cognition*, Cambridge: Massachusetts Institute of Technology Press, pp. 329-346.
RICHERSON, Peter, BOYD, Robert, 2005, *Not by Genes Alone; How Culture Transformed Human Evolution*, Chicago, London: University of Chicago Press.
RICHERSON, Peter, BOYD, Robert, HENRICH, Joseph, 2003, "Cultural Evolution of Human Cooperation", in P. HAMMERSTEIN (éd.), *Genetic and Cultural Evolution of Cooperation*, Cambridge: MIT Press, pp. 357-388.
RIDLEY, Matt, 2004 (2003), *Nature via Nurture; Genes, Experience, and what makes us human*, London: Harper Perennial.
RILLING, James, GUTMAN, David, ZEH, Thorsten, PAGNONI, Giuseppe, BERNS, Gregory, KILTS, Clinton, 2002, "Neural Basis for Social Cooperation", *Neuron*, 35, pp. 395-405.
RIOLO, Rick, COHEN, Michael, AXELROD, Robert, 2001, "Evolution of Cooperation without Reciprocity", *Nature*, 414, pp. 441-443.
ROBERTS, Gilbert, 1998, "Competitive Altruism; From Reciprocity to the Handicap Principle", *Proceedings of the Royal Society of London B*, 265, pp. 427-431.
ROBERTS, Gilbert, SHERRATT, Thomas, 1998, "Development of Cooperative Relationships Through Increasing Investment", *Nature*, 394, pp. 175-179.

ROBERTS, Simon, 1979, *Order and Dispute; an Introduction to Legal Anthropology*, New York: Penguin.
ROBINSON, Jenefer, 2004, "Emotion; Biological Fact or Social Construction?", in R. SOLOMON (éd.), *Thinking About Feeling; Contemporary Philosophers on Emotions*, Oxford, New York: Oxford University Press, pp. 28-43.
ROBINSON, Jenefer, 2005, *Deeper than Reason; Emotion and its Role in Literature, Music, and Art*, Oxford, New York: Oxford University Press.
ROGERS, Alan, 1988, "Does Biology Constrain Culture?", *American Anthropologist*, 90, pp. 819-831.
ROSENBERG, Alex, 1991, "The Biological Justification of Ethics", *Social Philosophy and Policy*, 8, pp. 86-101.
ROTTSCHAEFER, William, 1998 (1997), *The Biology and Psychology of Moral Agency*, Cambridge: Cambridge University Press.
ROTTSCHAEFER, William & MARTINSEN, David, 1990, "Really Taking Darwin Seriously; An Alternative to Michael Ruse's Darwinian Metaethics", *Biology and Philosophy*, 5, pp. 149-173.
ROUSSEAU, Jean-Jacques, 1999 (1762), *Emile ou de l'éducation*, Paris: Gallimard (folio essais).
ROUSSEAU, Jean-Jacques, 2001 (1762), *Du contrat social*, Paris: Garnier-Flammarion.
ROZIN, Paul, LOWERY, Laura, IMADA, Sumio, HAIDT, Jonathan, 1999, "The CAD Triad Hypothesis; A Mapping between Three Moral Emotions (Contempt, Anger, Disgust) and Threee Moral Codes (Community, Autonomy, Divinity)", *Journal of Personality and Social Psychology*, 76/4, pp. 574-586.
RUSE, Michael, 1984, "The Morality of the Gene", *Monist*, 67, pp. 176-199.
RUSE, Michael, 1986, "Evolutionary Ethics; A Phoenix Arisen", *Zygon*, 21, pp. 95-112.
RUSE, Michael (éd.), 1989, *Philosophy of Biology*, New York, London: Macmillan.
RUSE, Michael, 1993 (1991), "Une Défense de l'éthique évolutionniste", in J.-P. CHANGEUX *et al* (éds.), 1993 (1991), *Fondements naturels de l'éthique*, Paris: Odile Jacob, pp. 35-64.
RUSE, Michael, 1998 (1986), *Taking Darwin seriously; a naturalistic Approach to Philosophy*, New York: Prometheus Books.
RUSE, Michael, 2000, *The Evolution Wars; a Guide to the Controversies*, Santa Barbara: ABC-CLIO.
RUSE, Michael, 2002, "A Darwinian Naturalists Perspective on Altruism", in S. POST *et al.* (éds.), *Altruism and Altruistic Love*, Oxford, New York: Oxford University Press, pp. 151-167.
RUSE, Michael, WILSON, Edward, 1985, "The Evolution of Ethics", *New Scientist*, 17, pp. 50-52.
RUSE, Michael, WILSON, Edward, 1986, "Moral Philosophy as Applied Science", *Philosophy*, 61, pp. 173-192.
RYAN, James, 1997, "Taking the 'Error' Out of Ruse's Error Theory", *Biology and Philosophy*, 12, pp. 385-397.
SAGNER, Andreas, 2001, "The Abandoned Mother; Ageing, Old Age and Missionaries in Early and Mid Nineteenth-Century South-East Africa", *Journal of African History*, 42, pp. 173-198.
SAHLINS, Marshall, 1980 (1976), *Critique de la sociobiologie ; aspects anthropologiques*, trad. de l'angl. par J.-F. ROBERTS, Paris: Gallimard.
SANFEY, Alan, RILLING, James, ARONSON, Jessica, NYSTROM, Leigh, COHEN, Jonathan, 2003, "The Neural Basis of Economic Decision-Making in the Ultimatum Game", *Science*, 300, pp. 1755-1758.
SAYRE-MCCORD, Geoffrey (éd.), 1988, *Essays on Moral Realism*, Ithaca, London: Cornell University Press.
SAYRE-MCCORD, Geoffrey, 1997, " 'Good' On Twin Earth", *Philosophical Issues*, 8, pp. 267-292.
SAYRE-MCCORD, Geoffrey, 2005, "Moral Realism", in E. ZALTA (éd.), *The Stanford Encyclopedia of Philosophy*, edition d'hiver. Accessible en ligne:

<http://plato.stanford.edu/archives/win2005/entries/moral-realism/> (réf. 01.06.2006)
SCHELER, Max, 1955 (1913), *Le formalisme en éthique et l'éthique matériale des valeurs*, trad. de l'all. par M. de GANDILLAC, Paris: Gallimard.
SCHERER, Klaus, WALLBOTT, Harald, 1994, "Evidence for Universality and Cultural Variation of Differential Emotion Response Patterning", *Journal of Personality & Social Psychology* 66, pp. 310-328.
SCHRODINGER, Erwin, 1993 (1944), *Qu'est-ce que la vie ? de la physique à la biologie*, trad. de l'angl. par L. KEFFLER, Paris: Seuil.
SESARDIC, Neven, 1995, "Recent Work on human Altruism and Evolution", *Ethics*, 106/1, pp. 128-157.
SHERMAN, Paul, 1977, "Nepotism and the Evolution of Alarm Calls", *Science*, 197, pp. 1246-1253; in J. MAYNARD SMITH (éd.), 1982, *Evolution Now; a Century After Darwin*, London: Nature Publications, pp. 186-203.
SILK, Joan, 2003, "Cooperation without Counting; The Puzzle of Friendship", in P. HAMMERSTEIN (éd.), *Genetic and Cultural Evolution of Cooperation*, Cambridge: MIT Press, pp. 37-54.
SIMNER, Marvin, 1971, "Newborn's Response to the Cry of another Infant", *Developmental Psychology*, 5, pp. 136-150.
SINGER, Peter, 1981, *The Expanding Circle; Ethics and Sociobiology*, New York: Farrar, Straus & Giroux.
SMITH, Adam, 2003 (1759), *Théorie des sentiments moraux*, trad. de l'angl. Par M. BIZIOU *et al.*, Paris: PUF.
SMITH, Michael, 1986, "Should We Believe in Emotivism", in G. MACDONALD *et al.* (éds.), *Fact, Science, and Morality*, Oxford: Blackwell, pp. 289–310.
SOBER, Elliott, 1984a, "Force and Disposition in Evolutionary Theory", in C. HOOKWAY (éd.), *Minds, Machines and Evolution*, Cambridge: Cambridge University Press, pp. 43-62.
SOBER, Elliott, 1984b, *The Nature of Selection*, Cambridge (Mass.): Bradford Books/MIT Press.
SOBER, Elliott, 1992, "Hedonism and Butler's Stone", *Ethics*, 103/1, pp. 97-103.
SOBER, Elliott, 1993, "Evolutionary Altruism, Psychological Egoism and Morality; Disentangling the Phenotypes", in M. NITECKI *et al.* (éds.), *Evolutionary Ethics*, Albany: SUNY Press, pp. 199-216.
SOBER, Elliott (éd.), 1994a (1993), *Conceptual Issues in Evolutionary Biology*, Cambridge (Mass), London: MIT Press.
SOBER, Elliott, 1994b (1993), "Models of Cultural Evolution", in E. SOBER (éd.), 1994 (1993), *Conceptual Issues in Evolutionary Biology*, Cambridge (Mass), London: MIT Press, pp. 477-492.
SOBER, Elliott, 2007, "What is wrong with Intelligent Design? ", *The Quarterly Review of Biology*, 82/1, pp. 3-8.
SOBER, Elliott, WILSON, David S., 2002 (2000), "Are We Really Altruists?", in L. KATZ (éd.), 2002 (2000), *Evolutionary Origins of Morality; Cross-Disciplinary Perspectives*, in *Journal of Consciousness Studies*, 7, Bowling Green, Imprint Academic, pp. 185-206.
SOBER, Elliott, WILSON, David S., 2003 (1998), *Unto Others; The Evolution and Psychology of Unselfish Behavior*, London: Harvard University Press.
SOLTIS, Joseph, BOYD, Robert, RICHERSON, Peter, 1995, "Can Group-functional Behaviors Evolve by Cultural Group Selection?", *Current Anthropology*, 36/3, pp. 473-483.
SOMMERS, Tamler, ROSENBERG, Alex, 2003, "Darwin's Nihilistic Idea; Evolution and the Meaningless of Life", *Biology and Philosophy*, 18, pp. 653-668.
SPENCER, Herbert, 1879, *The Data of Ethics*, London: Williams and Norgate. Accessible en ligne: <http://fair-use.org/herbert-spencer/the-data-of-ethics> (réf. 21.05.2006)
SPENCER, Herbert, 1981 (1893), *Principles of Ethics*, Indianapolis: Liberty Fund.
SPERBER, Dan, 1996, *La contagion des idées*, Paris, Odile Jacob.
SPERBER, Dan, CLAIDIÈRE, Nicolas, 2008, "Defining and Explaining Culture", *Biology and Philosophy*, 23/2, pp. 283-292.

SPERBER, Dan, HIRSCHFELD, Lawrence, 2004, "The Cognitive Foundations of Cultural Stability and Diversity", *Trends in Cognitive Sciences*, 8/1, pp. 40-46.
SPINOZA, Baruch (de), 1965 (1661-1675), *Œuvres*, Vol. III : *Ethique ; démontrée suivant l'ordre géométrique et divisée en cinq parties*, trad. du latin par C. APPUHN, Paris: GF-Flammarion.
SRIPADA, Chandra, STICH, Stephen, 2005 (2004), "Evolution, Culture and the Irrationality of the Emotions", in D. EVANS *et al.* (éds.), 2005 (2004), *Emotion, Evolution and Rationality*. Oxford: Oxford University Press, pp. 133-158.
SRIPADA, Chandra, STICH, Stephen, 2006, "A Framework for the Psychology of Norms", in P. CARRUTHERS *et al.* (éds.), *The Innate Mind: Culture and Cognition*, Vol. II., pp. 280-301.
STEPHENS, David, MCLINN, Colleen, STEVENS, Jeffery, 2002, "Discounting and Reciprocity in an Iterated Prisoner's Dilemma", *Science*, 298, pp. 2216-2218.
STERELNY, Kim, 2006, "The Evolution and Evolvability of Culture", *Mind & Language*, 21/2, 2006, pp. 137-165.
STEVENSON, Charles, 1937, "The Emotive Meaning of Ethical Terms", *Mind*, 46, pp. 14-31.
STICH, Stephen, WEINBERG, Jonathan, 2001, "Jackson's Empirical Assumptions", *Philosophy and Phenomenological Research*, 62, pp. 637-643.
STURGEON, Nicolas, 1984, "Moral Explanations", in D. COPP *et al.* (éds.), *Morality, Reason and Truth*, Totowa: Rowman & Allanheld, pp. 49-78.
SUGDEN, Robert, 1986, *The Economics of Rights, Co-operation and Welfare*, Oxford: Blackwell.
SUNSTEIN, Cass, 2005, "Moral Heuristics", *Behavioral and Brain Sciences*, 28, pp. 531-542.
TAPPOLET, Christine, 2000, *Emotions et valeurs*, Paris: PUF.
TAYLOR, Peter, 1992, "Altruism in Viscous Populations - An Inclusive Fitness Approach", *Evolutionary Ecology*, 6, pp. 352-6.
THAGARD, Paul, 1992, *Conceptual Revolutions*, Princeton: Princeton University Press.
THOMPSON, Paul, 1999, "Evolutionary Ethics; its Origins and Contemporary Face", *Zygon*, 34/3, pp. 473-484.
TOMASELLO, Michael, 2004 (1999), *Aux origines de la cognition humaine*, trad. de l'angl. par Y. BONIN, Paris: Retz (Forum éducation culture).
TOMASELLO, Michael, CALL, Joseph, HARE, John, 2003, "Chimpanzees Understand Psychological States; The Question is which Ones and to what Extend", *Trends in Cognitive Sciences*, 7/4, pp. 153-156.
TOMASELLO, Michael, CARPENTER, Malinda, CALL, Joseph, BEHNE, Tanya, MOLL, Henrike, 2005, "Understanding and Sharing Intentions; The Origins of Cultural Cognition", *Behavioral and Brain Sciences*, 28, pp. 675–735.
TOMASELLO, Michael, KRUGER, Ann, RATNER, Hillary, 1993, "Cultural learning", *Behavioral and Brain Sciences*, 16, pp. 495-552.
TOOBY, John, COSMIDES, Leda, 1989, "Evolutionary Psychologists Need to Distinguish between the Evolutionary Process, Ancestral Selection Pressures, and Psychological Mechanisms", *Behavioral and Brain Sciences*, 12, pp. 724-725.
TORT, Patrick, 2002, *La seconde révolution darwinienne ; biologie évolutive et théorie de la civilisation*, Paris: Kimé.
TRIVERS, Robert, 1971, "The Evolution of Reciprocal Altruism", *The Quarterly Review of Biology*, 46, 1, pp. 35-57.
TRIVERS, Robert, 1985, *Social Evolution*, Menlo Park: Benjamin Cummings.
TURIEL, Elliot, 1983, *The Development Of Social Knowledge; Morality and Convention*, Cambridge: Cambridge University Press.
TURIEL, Elliot, 1993 (1991), "Nature et fondements du raisonnement social dans l'enfance", in J.-P. CHANGEUX *et al.* (éds.), 1993 (1991), *Fondements naturels de l'éthique*, Paris: Odile Jacob, pp. 301-317.
VIRVIDAKIS, Stélios, 1996, *La robustesse du Bien; Essai sur le Réalisme moral*, Nîmes: Jacqueline Chambon.
VOORZANGER, Bart, 1987, "No Norms and no Nature; The Moral Relevance of Evolutionary

Biology", *Biology and Philosophy*, 3, pp. 253-270.
VON NEUMANN, John, 1928, "Zur Theorie der Gesellschaftsspiele", in *Mathematische Annalen*, 100, Berlin: Springer, pp. 295-320.
VON NEUMANN, John, MORGENSTERN, Oskar, 1944, *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.
VON WRIGHT, Georg, 1963, *Norm and Action*, London: Routledge.
WALLER, Bruce, 1997, "What Rationality Adds to Animal Morality", *Biology and Philosophy*, 12, pp. 341-356.
WEBER, Bruce, DEPEW, David (éds.), 2003, *Evolution and Learning; The Baldwin Effect Reconsidered*, Cambridge: MIT Press.
WEDEKIND, Claus, BRAITHWAITE, Victoria, 2002, "The Long-Term Benefits of Human Generosity in Indirect Reciprocity", *Current Biology*, 12, pp. 1012–1015.
WEDEKIND, Claus, MILINSKI, Manfred, 2000, "Cooperation Through Image Scoring in Humans", *Science*, 288, pp. 850-852.
WEGNER, Daniel, 2002, *The Illusion of Conscious Will*, Cambridge (Mass.): MIT Press.
WEST, Stuart, MURRAY, Martyn, MACHADO, Carlos, GRIFFIN, Ashleigh, HERRE, Edward, 2001, "Testing Hamilton's Rule with Competition between Relatives", *Nature*, 409, pp. 510-513.
WHEATLEY, Thalia, HAIDT, Jonathan, 2005, "Hypnotic Disgust Makes Moral Judgments More Severe", Psychological Science, 16/10, pp. 780-784.
WILBERFORCE, Samuel, 1860 (publié anonymement), "Is Mr Darwin a Christian? Review of On The Origin of Species", Quarterly Review, 108, pp. 225-264; in WILBERFORCE, Samuel, 1874, *Essays Contributed to the Quarterly Review*, Vol.1, pp. 52-103.
WILKINSON, Gerald, 1990, "Le partage du sang chez les vampires", trad. de l'angl., *Pour la Science*, 150, pp. 58-65.
WILKINSON, Gerald, 1984, "Reciprocal Food Sharing in the Vampire Bat", *Nature*, 308, pp. 181-184.
WILLIAMS, Bernard, 1973, "Morality and the Emotions", in *Problems of the Self; Philosophical Papers 1956-1972*, Cambridge: Cambridge University Press, pp. 207-229.
WILLIAMS, Bernard, 1985, *Ethics and the Limits of Philosophy*, Cambridge (Mass.): Harvard University Press.
WILLIAMS, George, 1966, *Adaptation and Natural Selection; A Critique of Some Current Evolutionary Thought*, Princeton: Princeton University Press.
WILLIAMS, George, 1988, "Huxley's Evolution and Ethics in Sociological Perspective", *Zygon*, 23, pp. 383-407.
WILLIAMS, George, 1993, "Mother Nature Is a Wicked Old Witch", in M. NITECKI *et al.* (éds.), *Evolutionary Ethics*, Albany: SUNY Press, pp. 217-231.
WILLIAMS, Patricia, 1990, "Evolved Ethics Re-Examined The Theory of Robert J. Richards", *Biology and Philosophy*, 5, pp. 451-457.
WILLIAMS, Patricia, 1993, "Can Beings Whose Ethics Evolved Be Ethical Beings? ", in H. NITECKI *et al.* (éds.), *Evolutionary Ethics*, Albany: SUNY Press, pp. 217-239.
WILSON, David S., 1975, "A General Theory of Group Selection", *Proceedings of the National Academy of Sciences*, 72, pp. 143-146.
WILSON, David S., 2002, "Evolution, Morality and Human Potential", in S. SCHER *et al.* (éds.), *Evolutionary Psychology; Alternative Approaches*, Boston, Dordrecht, New York, London: Kluwer Academic Publishers, pp. 55-70.
WILSON, Edward O., 1975, *Sociobiology; The New Synthesis*, Cambridge: Harvard University Press.
WILSON, Edward O., 1979 (1978), *L'humaine Nature*, trad. de l'angl. par Roland BAUCHOT, Paris: Stock.
WOOLKOCK, Peter, 1993, "Ruse's Darwinian Meta-Ethics; A Critique", *Biology and Philosophy*, 8, pp. 423-439.
WRIGHT, Larry, 1973, "Functions", *The Philosophical Review*, 82/2, pp. 139-16; in E. SOBER (éd.), 1994 (1993), *Conceptual Issues in Evolutionary Biology*, Cambridge (Mass), London: MIT Press, pp. 27-48.

WRIGHT, Robert, 1994 (1995), *L'Animal moral; psychologie évolutionniste et vie quotidienne*, trad. de l'angl. par A. BÉRAUD-BUTCHER, Paris: Editions Michalon.
WRIGHT, Sewall, 1931, "Evolution in Mendelian Populations", *Genetics*, 16, pp. 97-159.
WRIGHT, Sewall, 1932, "The Role of Mutation, Inbreeding, Crossbreeding and Selection in Evolution", *Proceedings of the 6th International Congress of Genetics*, 1, pp. 356-368.
WRIGHT, Sewall, 1945, "Tempo and Mode in Evolution; A Critical Review, Ecology", 26, pp. 415-419.
WYNNE-EDWARDS, Vero, 1962, *Animal Dispersion in Relation to Social Behaviour*, Edinburgh: Olivier and Boyd Publishing.
XENOPHANE, 1988 (5ème s. av. J.-C.), "B. I; Fragments", in J.-P. DUMONT *et al.* (éds.), *Les présocratiques*, Paris: Gallimard (La Pléiade), pp. 113-124.
YABLO, Stephen, 2000, "Red, Bitter, Best," *Philosophical Books*, 41, pp. 13-23.
ZAHAVI, Amotz, 1975, "Mate Selection – A Selection for a Handicap", *Journal of Theoretical Biology*, 53/1, pp. 205-214.
ZAHAVI, Amotz, 1977, "Reliability in Communication Systems and the Evolution of Altruism", in B. STONEHOUSE *et al.* (éds.), *Evolutionary Ecology*, London: Macmillan Press, pp. 253-259.
ZAHAVI, Amotz, 2002 (2000), "Altruism; The Unrecognized Selfish Traits", in L. KATZ (éd.), 2002 (2000), *Evolutionary Origins of Morality; Cross-Disciplinary Perspectives*, in *Journal of Consciousness Studies*, 7, Bowling Green, Imprint Academic, pp. 253-256.
ZENTALL, Thomas, 2006, "Imitation: definitions, evidence, and mechanisms", *Animal Cognition*, 9, pp. 335-353.
ZIMMERMAN, David, 1984, "Moral Realism and Explanatory Necessity", in D. COPP *et al.* (éds.), *Morality, Reason and Truth*, Totowa: Rowman & Allanheld, pp. 79-103.

Index des noms propres

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

V

W

Z