

Introduction

Deuxième partie : Petite introduction aux théories évolutionnistes

In : Clavien, C. & El-Bez, C. (eds.), *Morale et évolution biologique ; entre déterminisme et liberté*, Lausanne : PPUR, sous presses (2007)

Christine Clavien

Université de Neuchâtel
Institut de philosophie
Espace Louis-Agassiz 1
CH-2001 Neuchâtel
christine.clavien@unine.ch

1 Petit bréviaire de la théorie de l'évolution

D'ordinaire, lorsqu'il est question de théorie de l'évolution, on pense à la théorie de l'évolution des espèces biologiques dont on doit la première formulation 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 pas 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 publiait 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 sont les suivantes : Au cours du temps les espèces (des ensembles d'individus très semblables du point de vue de leurs caractéristiques observables) changent et descendent les unes des autres par le biais des voies ordinaires de la génération. Les espèces peuvent subsister pour une certaine période, disparaître ou se diversifier au fil des générations, donnant éventuellement naissance à de nouvelles espèces. Ainsi, 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 : 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, il s'agit du plus important mécanisme de tri parmi la diversité.

La sélection naturelle s'applique en principe à toute population d'entités qui remplissent simultanément trois conditions : variation (entre les différentes entités sujettes à la sélection), reproduction (des entités sujettes à la sélection) et hérédité (des caractéristiques propres à la génération précédente).

Considérons le mécanisme de la sélection naturelle appliqué à des organismes vivants : Pour qu'il y ait sélection, ces organismes doivent être variés du point de vue de leurs traits (caractéristiques observables) et de leur *fitness* (valeur de survie et de reproduction) et pouvoir transmettre leurs traits aux générations suivantes (via la reproduction).

Dans les écrits contemporains, on trouve différentes définitions de la *fitness*. Les conceptions les plus connues sont celles de la *fitness classique* et de la *fitness inclusive*. Considérons pour le moment uniquement la première. Il s'agit d'une propriété d'un individu qui comprend deux volets : 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. On dit d'organismes dont la viabilité et/ou la fécondité sont différentes que leurs *fitness* sont différentes. 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. Dans ce contexte, on parle aussi d'adaptation : si un individu a une plus grande *fitness* qu'un de ces congénères, on dira du premier qu'il est mieux adapté que le second à l'environnement dans lequel il évolue.¹

Revenons à la sélection naturelle : il s'agit du processus durant lequel les organismes dotés d'une meilleure *fitness* (c'est-à-dire les individus les mieux adaptés) 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 (ou plus précisément le matériel héréditaire qui leur est sous-jacent) ; 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

¹ La notion d'adaptation est également intimement liée à celle de fonction biologique (pour une bonne définition de la fonction, voir DE SOUSA ce volume, p.--).

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.

Trois 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, d'autres traits seront sélectionnés. Troisièmement, 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 dotés d'une meilleure *fitness* 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.²

2 *Le paradoxe de l'altruisme*

Un certain nombre de biologistes (à commencer par DARWIN) se sont intéressés à l'évolution du comportement social chez les organismes vivants. Dans le cadre de ces recherches, ils se sont rapidement trouvés confrontés à un paradoxe désormais devenu fameux : le paradoxe de l'altruisme.

Pour comprendre ce paradoxe, commençons par définir ce que les biologistes entendent par « altruisme ». Ils conçoivent cette notion en termes de *fitness*. Plus précisément : *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.*

Quelques éclaircissements terminologiques s'imposent ici : Tout d'abord, dans le contexte des théories évolutionnistes, lorsque l'on parle d'un individu altruiste, on pense à un

² 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).

individu dont la base génétique induit une tendance plus ou moins forte à adopter un comportement altruiste.

D'autre part, la notion d'altruisme évolutionniste 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 évolutionniste n'entretient aucun rapport logique avec les motifs d'action des individus ; l'altruisme évolutionniste se calcule uniquement en fonction des conséquences des comportements sur la *fitness* des organismes. Cette notion se distingue donc nettement de la conception ordinaire que nous nous faisons de l'altruisme (cette distinction sera élaborée dans la dernière section).

Après ces éclaircissements, revenons à la définition du comportement altruiste et considérons un exemple illustratif : Selon la définition de la *fitness classique* donnée plus haut, la *fitness* d'un individu se calcule en termes de viabilité (capacité à survivre) et de fécondité (capacité à 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 ; en effet, si l'abeille pique, elle voit sa viabilité réduite à néant tout en augmentant les chances de survie de ses congénères.

Les traits véritablement altruistes dans la nature posent un défi aux biologistes car ils semblent se soustraire à la logique de sa théorie de la sélection naturelle ; en effet, on voit mal comment la théorie de la sélection naturelle a pu sélectionner un tel trait qui aurait dû être éliminé au même titre que tous les autres traits nuisibles aux individus qui les possèdent.

Voyons dans le détail en quoi consiste ce paradoxe de l'altruisme (le lecteur peu intéressé aux détails peut passer à la section suivante) :

Prenons 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 potentielle 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 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 ceci reste valable quelle que soit la proportion d'altruistes au sein du groupe.³ 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 amènent un bénéfice global en terme 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 montre 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 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 les 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 une invention ; il s'agit d'une observation empirique devant laquelle Darwin lui-même est resté perplexe. Ainsi, les théoriciens de

³ 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 se trouvent devant un problème dérangentant ; leur fameuse théorie ne permet pas d'expliquer un type de phénomènes empiriquement observable. Depuis DARWIN, un grand nombre de penseurs ont tenté de faire disparaître le paradoxe de l'altruisme. Retraçons les différentes explications qui ont été proposées.

3 Darwin et la sélection de groupe

Pour résoudre le paradoxe de l'altruisme et expliquer des comportements tels que ceux des abeilles kamikazes, DARWIN a proposé une théorie de sélection de groupe (2000 (1871)). Cette théorie considère le groupe comme une unité de sélection ; ainsi, entre différents groupes donnés, au cours du processus de sélection naturelle, il y a survie de certains et disparition des autres (en d'autres termes, les groupes sont en compétition constante dans la nature). Or, il se trouve que beaucoup de comportements altruistes, bien que nuisibles aux individus qui les exercent, ne sont pas pour autant inutiles au niveau du groupe ; ils pourraient même lui permettre de s'imposer face à d'autres groupes qui ne sont pas composés d'individus altruistes. Ainsi, il est tout à fait possible d'imaginer qu'un trait bénéfique pour un groupe peut être sélectionné en dépit du coût qu'il engendre aux individus qui le portent ; en d'autres termes, 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 ne comptent que des non-altruistes, ils sélectionneront dans leur sillage, leurs membres altruistes

Cette théorie a connu de beaux jours jusque dans les années 60 où des biologistes ont mis en question l'idée qu'un groupe puisse servir d'unité de sélection. L'argument massue qui semble ruiner la théorie de la sélection de groupe est le suivant :

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)

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 que cette 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.

En concurrence de la sélection de groupe une nouvelle théorie, nettement plus prometteuse, a été proposée par William HAMILTON (1964) : la sélection de parentèle. Pour la comprendre, il vaut la peine d'esquisser le contexte intellectuel dans lequel elle est apparue.

4 La perspective du gène et la théorie de la sélection de parentèle

a. La perspective du gène et la fitness inclusive

Depuis le début du vingtième siècle, grâce à la redécouverte (DE VRIES 1900) des travaux de MENDEL (1913) sur l'hérédité, 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. 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.

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 ». ⁴ C'est depuis lors que l'on a commencé à adopter la perspective du gène, ⁵ 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. 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 ⁶ 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

⁴ Cette formule a été rendue célèbre par Richard DAWKINS (1996 (1976)).

⁵ La perspective du gène a été diffusée par un courant appelé « sociobiologie » dont les défenseurs se sont exprimés dans les années 60 et 70 ; les sociobiologistes tentent de rendre compte du comportement social des êtres vivants (homo sapiens y compris). Ce courant de pensée ayant été l'objet d'âpres débats en partie politisés, de nos jours, les biologistes de l'évolution intéressés au comportement social préfèrent ne plus utiliser ce terme pour désigner leurs théories.

⁶ Le pool génétique d'une espèce (ou population) est l'ensemble des gènes (plus particulièrement des allèles) véhiculés par les individus vivants de cette espèce (ou population).

dans la mesure où elle sert ‘l’intérêt’⁷ des gènes véhiculés par cet individu. 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 à 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).

La *fitness inclusive* est une mesure qui cumule à la fois la viabilité et la fécondité individuelles et les effets du comportement d’un individu sur la viabilité et la fécondité des individus de son voisinage social (lesquels sont généralement des individus parents).

b. La théorie de la sélection de parentèle

L’importance de cette nouvelle définition est qu’elle permet de résoudre le paradoxe de l’altruisme devant lequel bute une conception de la sélection naturelle en termes de *fitness classique*. En effet, si au lieu de se demander comment un individu peut transmettre un trait comportemental altruiste à ses descendants, on se demande comment les gènes qui induisent ce type de comportements peuvent se répandre dans l’ensemble du pool génétique d’une population, on trouve la clef de la solution : les gènes peuvent se répandre dans le pool génétique d’une population à la condition qu’ils induisent, chez l’individu porteur de ces gènes (l’altruiste), des comportements altruistes en faveur des individus porteurs de copies d’eux-mêmes (en général ce sont les proches parents de l’altruiste). En d’autres termes, 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.

Voyons ce qu’il en est dans le détail (le lecteur peu intéressé aux détails peut passer à la section suivante). Pour savoir si la valeur de la *fitness inclusive* est positive ou négative, on procède à un calcul assez complexe qui comporte deux étapes : premièrement le calcul de l’apparentement génétique (r), ensuite l’application de la règle de Hamilton. Considérons une étape après l’autre :

HAMILTON a établi une manière de calculer le coefficient d’apparentement entre deux individus (r). Sans entrer dans des formules mathématiques complexes, tâchons d’en rendre

⁷ Il est clair que lorsque l’on parle de l’‘intérêt’ des gènes, c’est dans un sens métaphorique (au même titre que la notion de ‘gène égoïste’ utilisée par bon nombre d’auteurs) ; les gènes n’ont évidemment pas d’intentions.

compte à l'aide d'un modèle de groupe : r est une notion relative qui établit une relation entre deux individus issus d'un même voisinage social, relativement à une population de référence. Je m'explique : Le voisinage social est une notion souple qui peut être comprise comme un groupe d'individus qui interagissent régulièrement (par exemple une tribu très soudée) ou plus simplement comme un nombre restreint d'individus parents (par exemple des frères et sœurs issus des deux mêmes parents). La population de référence correspond au domaine de compétition, c'est-à-dire à un ensemble d'individus d'une même espèce, susceptibles d'entrer en compétition (par exemples, tous les individus qui vivent sur une île aux ressources naturelles limitées). Dans ce tableau, r correspond à la proportion attendue de gènes communs entre un individu X et n'importe quel individu Y issu du même voisinage social (par exemple entre X et son frère Y)⁸, en excès de la proportion moyenne de gènes communs entre X et un individu quelconque de la population de référence (par exemple entre X et un individu pris au hasard dans l'île). Et si l'on adopte une perspective géno-centrique, r correspond à la probabilité qu'un gène d'un individu X se trouve également dans n'importe quel individu Y issu du même voisinage social (si le gène est rare, entre X et son frère, la probabilité sera d'environ 50%), en excès de la probabilité qu'il se trouve dans un individu quelconque de la population de référence (si le gène est rare, entre X et un individu pris au hasard dans l'île, la probabilité sera très faible).

Venons-en à la règle de Hamilton. Cette règle dit que, considérant un comportement (en l'occurrence, un comportement altruiste), si la *fitness inclusive* relative à ce comportement est positive, alors le comportement est susceptible d'évoluer. Plus précisément, si la *fitness inclusive* est positive, cela signifie que les gènes qui induisent le comportement coopératif ou altruiste ont pu être transmis à la génération suivante. Mathématiquement parlant, la *fitness inclusive* est positive lorsque

$$rb > c$$

où :

r = le coefficient d'apparentement entre le donneur (celui qui produit une action altruiste) et le/les receveur/s (celui/ceux qui en bénéficie/nt)

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

⁸ En général, si le voisinage social est constitué uniquement des deux individus, cette proportion correspond au degré de parenté par voie de descendance. 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.

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

En d'autres termes, pour qu'un comportement altruiste génétiquement déterminé puisse évoluer, il faut que le bénéfice (du/des receveur/s) multiplié par le coefficient d'apparentement soit supérieur au coût (du donneur). Cette règle signifie simplement qu'un comportement évolue si, par ce biais, les gènes responsables de ce comportement laissent au total un plus grand nombre de copies.

Ainsi, la sélection des gènes responsables de l'altruisme est fortement dépendante du type de bénéficiaires des actions altruistes : il faut qu'un nombre significatif 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 s'il possède à la fois la capacité à reconnaître ses proches parents et celle d'agir de manière discriminatoire en leur faveur. 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 (un parent donne 50% de ses gènes à son enfant). Ainsi, la règle de Hamilton prédit qu'un comportement altruiste envers des frères sera sélectionné uniquement s'il occasionne un gain (pour les frères) plus de deux fois supérieur à la perte occasionnée à l'agent (puisque deux frères partagent 50% de leurs gènes) ; en faveur de demi-frères, le gain devra encore doubler etc.

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 (a) 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 ou (b) s'il dispense son altruisme de manière non discriminatoire dans l'environnement qu'il habite *et* que cet environnement est essentiellement composé d'autres individus altruistes.

c. La résolution du mystère des abeilles kamikazes

Pour terminer, illustrons cette théorie de la sélection de parentèle par un exemple concret : Dans son article de 1964, HAMILTON parvient à résoudre le mystère des

comportements kamikazes des abeilles. Pour commencer, il faut savoir que la majorité des individus composant une colonie d'abeilles sont des ouvrières stériles. Dès lors, il n'est plus guère étonnant que celles-ci se dévouent à leur communauté ; les comportements altruistes tels que les attaques 'suicidaires' des abeilles trouvent une explication assez directe. Un problème demeure toutefois : Comment se fait-il que ce phénomène (l'existence d'ouvrières stériles) ait pu être sélectionné au fil de l'évolution ? La stérilité n'est-elle pas un facteur d'extinction ? En réalité, pour répondre à cette question, il suffit d'appliquer la règle de Hamilton : Chez les hyménoptères (ordre d'insectes comprenant les fourmis, abeilles, guêpes), les mâles naissent d'œufs non fécondés ; ils n'ont pas de père et toutes les cellules de leurs corps ne contiennent qu'un seul ensemble de chromosomes (tous proviennent de la mère) ; techniquement parlant, on dit de ces mâles qu'ils sont haploïdes. Les femelles par contre, sont normales ; elles naissent d'œufs fécondés et possèdent deux ensembles de chromosomes (un ensemble provenant de la mère et un ensemble provenant du père) ; elles sont donc diploïdes. De ce phénomène découle un tableau des relations de parenté bien différent de celui des autres animaux (et des êtres humains).

B	Mère	Père	Soeur	Frère	Fils	Fille
A						
Femelle	50%	50%	75%	25%	50%	50%
Mâle	100%	0%	50%	50%	0%	100%

Tableau inspiré de KREBS & DAVIES 1993 (1981), p. 330.

A lire dans le sens suivant : « pourcentage de gènes que A partage avec B = ...% » (ex : le pourcentage de gènes qu'un mâle partage avec son père est de 0%)

Il s'ensuit qu'une femelle hyménoptère possède un pourcentage de gènes communs plus élevé avec ses sœurs qu'avec sa progéniture (quel qu'en soit le sexe). Ainsi, il est plus intéressant pour une femelle de favoriser la production de ses sœurs plutôt que d'engendrer sa propre descendance. Dès lors, on comprend mieux comment des comportements kamikazes ont pu être sélectionnés chez les hyménoptères.

Ainsi, grâce à la théorie de HAMILTON, on peut expliquer 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 ; c'est le cas s'ils induisent des

comportements de sacrifice au profit des individus qui ont toutes les chances de posséder des copies d'eux-mêmes ; en d'autres termes, cela est possible si les comportements qu'ils induisent favorisent leur réplication.

d. La sélection de groupe se réduit à la sélection de parentèle

Comme on l'a vu, la théorie de la sélection de groupe, qui a été énoncée pour la première fois par Charles DARWIN pour expliquer l'évolution des comportements altruistes observés dans la nature, a été mise en doute par les partisans de la perspective du gène. Malgré cela, elle n'a pas perdu tous ces adeptes et il faut y revenir brièvement car beaucoup d'encre a coulé à son propos ces dernières années. Un des défenseurs les plus connus de la sélection de groupe, David Sloan WILSON (1975 ; WILSON & SOBER 2003), a proposé un modèle mathématique censé montrer que, dans certaines circonstances, la force de la sélection de groupe (qui favoriserait l'altruisme) peut supplanter la force de sélection à l'intérieur des groupes (qui opère au détriment de l'altruisme).

Pourtant, il est possible de montrer qu'en faisant abstraction de la parure théorique proposée par WILSON, sa théorie de la sélection de groupe se réduit ni plus ni moins à la règle de Hamilton ! (MAYNARD SMITH 1976). En d'autres termes, 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 fait encore sens de dire avec WILSON que la sélection agit à plusieurs niveaux, c'est-à-dire qu'elle exerce une *force* au niveau des individus et une *force* au niveau des groupes. MAYNARD SMITH et bien d'autres biologistes dans son sillage (GILDENHUYS 2003 ; PERRIN ce volume ; Laurent KELLER, communication personnelle) ne le pensent pas. Pour eux, ce que propose 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. Et s'il fait sens de parler d'une dynamique de groupe, il faut réfléchir en termes de voisinage social et domaine de compétition tels qu'ils ont été définis dans le chapitre sur la sélection de parentèle.

5 Les théories de la réciprocité

Nous avons vu que la théorie de la sélection de parentèle permet de résoudre le paradoxe de l'altruisme. Toutefois, cette théorie a ses limites. Elle peut uniquement expliquer les comportements altruistes entre individus parents (ou plus rarement, entre individus qui partagent le gène responsable de l'altruisme). Or, dans le monde animal et humain, on peut également observer des comportements altruistes entre individus non parents, voire même entre individus d'espèces différentes. Par exemple, l'éthologue Gérard WILKINSON (1984) et ses collaborateurs ont mené un travail d'observation sur le comportement d'une espèce de chauves-souris vampires (les *Desmodus rotundus*) et ont pu constater que les individus qui rentrent repus après une nuit de chasse régurgitent souvent une partie de leur repas dans la gueule des malchanceux rentrés bredouilles, ces derniers n'étant pas systématiquement des individus parents. Mais les exemples les plus frappants d'altruisme envers des étrangers peuvent évidemment être trouvés chez les êtres humains.

Pour expliquer ce genre de phénomènes, une nouvelle théorie a été développée : celle de la réciprocité directe.

a. 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 la 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 la 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 équivalent. Cela est possible s'ils possèdent la capacité à reconnaître et à se souvenir des individus envers lesquels ils ont agi de manière altruiste 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 (gain et coût doivent être pensés en termes de *fitness*) ; 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 équivalents pour les différents protagonistes. Considérons par exemple deux individus d'une troupe de singes très hiérarchisée ; le premier est un chef de troupe qui bénéficie de tous les privilèges, le second un jeune singe freluquet au bas de la hiérarchie. Dans ce cas, les deux individus ne sont pas exposés de manière symétrique aux interactions altruistes (puisque le chef de troupe peut de toute façon se servir de tous les biens qu'il convoite).

b. La théorie des jeux⁹ au service de l'altruisme réciproque – L'exemple de l'épouillage mutuel

⁹ A l'aide d'outils mathématiques et informatiques, la théorie des jeux cherche à simuler des situations d'interaction afin de déterminer quels comportements sont les plus appropriés dans des situations données. Elle a surtout été utilisée par les économistes pour comprendre les mécanismes de la société et prédire ce qui se produit dans différentes situations économiques. Mais comme on le verra, on peut également y recourir pour mieux comprendre des phénomènes sociaux comme la coopération ou l'altruisme.

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 mathématiciens (plus précisément des théoriciens des jeux) comme John MAYNARD SMITH (1982) ou Robert AXELROD (1996 (1984)) 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. Il s'agissait non seulement de tester la robustesse et la stabilité de différentes stratégies 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. 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 parasites affectionnent tout particulièrement ; la seule manière de se débarrasser de ces récalcitrantes est de se laisser épouiller par un autre oiseau, mais cela implique un investissement en temps et en énergie pour l'épouilleur. Supposons également que toutes les conditions nécessaires à la réciprocité listées ci-dessus soient réalisées (capacité de reconnaissance, occurrences régulières des situations d'interaction, inversion des rôles, rapport gain/coût et symétrie de l'interaction).

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 paye aucun coût en échange. En simulant l'évolution probable de ce type d'interactions, on obtient le résultat suivant : au fil des générations, la stratégie comportementale E va disparaître sous la pression de la sélection naturelle.¹⁰ 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, à E et T, 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 le mathématicien Robert AXELROD qui a découvert les vertus de cette stratégie altruiste réciproque. Je reviendrai plus loin sur ses

¹⁰ 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 puisque les oiseaux seront infestés de parasites dont ils ne pourront pas se débarrasser et qui causeront leur mort).

recherches mais avant cela, j'aimerais présenter une notion très commode lorsque l'on étudie l'évolution de stratégies comportementales : celle de stratégie évolutionnairement stable.

c. Stratégie évolutionnairement stable

La notion de stratégie évolutionnairement stable (SES) a été élaborée par John MAYNARD SMITH (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, 2/3 des individus d'une population pratiquent la stratégie A et 1/3 la stratégie B) ; dans ce cas on parle de stratégie mixte. En guise d'illustration, reprenons l'exemple de l'épouillage mutuel :

Si, dans le milieu de départ, on trouve une répartition des stratégies T (Tricheur), E (Epouilleur) et R (Rancunier) telle qu'il existe beaucoup de T contre 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 ce 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 parasites !).

Si, dans le milieu de départ, on trouve une répartition des stratégies T, E et R telle que E et R soient aussi bien représentées que T, au fil des générations, la sélection 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, par contre 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 pour la première fois ; dans ces cas, ils sont servis ! Ainsi, on se retrouve avec une stratégie mixte évolutionnairement stable (composée d'une grande majorité de Rancuniers et d'une petite proportion de Tricheurs).

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 puisqu'elle ne peut pas être supplantée par une stratégie existant dans l'environnement dans lequel elle s'est stabilisée. Une autre vertu de la SES est qu'à l'équilibre, toutes les stratégies présentes dans l'environnement ont la même *fitness*. Par contre, la perfection de la SES est limitée en trois

sens : 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 on a vu que si T devient une SES, les individus seront affaiblis par la maladie pour cause de manque de soins). Deuxièmement, une SES peut être déstabilisée par l'apparition d'une nouvelle stratégie mutante. 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.

d. Le dilemme du prisonnier réitératif

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 réitératif et un grand nombre de stratégies utilisables dans ce jeu. Comme nous le verrons, l'intérêt du dilemme du prisonnier réitératif 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 :

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 réitératif 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 défini 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 :

		Joueur B	
		<i>coopération</i>	<i>défection</i>
Joueur A	<i>coopération</i>	3 / 3	0 / 5

<i>défection</i>	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 une 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 « choisi 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 première rencontre ! » b/ « copie l'action précédente de ton adversaire ! » ; en d'autres termes, s'il a coopéré la dernière fois que tu l'as rencontré, coopère et inversement (en réalité, il s'agit de la stratégie du rancunier vue précédemment).

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 être une bonne stratégie tient au fait qu'elle est utilisée dans un jeu réitératif. Le fait qu'un grand nombre de coups soient joués et que les différents gains s'additionnent change 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 possible au fil des rencontres. Dans ces conditions, les interactions coopératives portent leurs fruits (trois points à chaque rencontre). Ainsi, une stratégie coopérative s'en sort bien 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.¹¹ 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.¹² 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 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 dans le tournoi suivant (en termes de nombre joueurs qui appliquent ces stratégies) 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 d'individus coopérateurs soient intégrés pour qu'au fil des tournois, *Donnant Donnant* évince les autres stratégies jusqu'à atteindre une stabilité évolutionnaire.¹³ Notons toutefois que ces résultats n'indiquent rien sur l'origine évolutionnaire 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).

Si on applique la théorie d'AXELROD au monde biologique, on admet que sous-jacent aux stratégies comportementales, il y a des gènes ; ainsi, la stratégie serait un type de

¹¹ 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, p. 129).

¹² Notons également que l'exemple des épouilleurs est une forme de dilemme du prisonnier.

¹³ Toutefois, dans un monde de défection inconditionnelle, quelques individus adoptant la stratégie *Donnant Donnant* ne pourront prospérer parce qu'ils ne rencontreront pas suffisamment de partenaires coopératifs. D'autre part, 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).

comportement (plus précisément un effet phénotypique) résultant d'une base génétique. 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 potentiel de la coopération ne soit pas trop élevé¹⁴ ; 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 devenir évolutionnairement stable, AXELROD fait appel à HAMILTON. Selon eux, « 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, 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 bien¹⁵ et générera une ligne de descendants qui peu à peu, pourront s'imposer au-delà de ce contexte restreint.

e. Les nouvelles théories de la réciprocité

Les travaux de MAYNARD-SMITH et AXELROD en théorie des jeux évolutionniste 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 réitératif développé par AXELROD en modifiant les conditions de jeu du dilemme du prisonnier de manière à le rendre plus proche de la réalité (par exemple ROBERTS & SHERRATT 1998).

¹⁴ 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 être une stratégie efficace étant donné le trop grand risque lié à la coopération.

¹⁵ A ce propos, précisons que l'altruisme réciproque peut aussi bien être pratiqué entre individus parents qu'entre individus non parents.

D'autres chercheurs ont tenté de proposer des alternatives aux modèles d'altruisme réciproque. Ronald NOË et Peter HAMMERSTEIN (1994) par exemple, partant de la constatation qu'il est extrêmement difficile de trouver des cas d'altruisme réciproque sur le modèle Donnant Donnant dans le monde animal,¹⁶ ont développé la « théorie du marché biologique » (*biological market theory*) qui se veut plus réaliste en intégrant le paramètre du choix des partenaires de coopération. Selon ces auteurs, 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 réitératifs). 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 retourner 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, plutôt que sous forme de Donnant Donnant,¹⁷ 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.

f. La réciprocité n'est pas altruiste

Au terme de cette analyse de la réciprocité, une question importante se pose : Dans un 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* d'autrui aux dépens de la *fitness* (au sens classique du terme) 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, il faut admettre que les cas d'altruisme réciproque n'entrent pas sous la définition de l'altruisme évolutionniste. A ce propos, on peut même dire

¹⁶ Il n'est par exemple pas évident que les comportements des chauves-souris observées par WILKINSON puissent être assimilés à la stratégie Donnant Donnant (à ce propos, voir HAMMERSTEIN 2003).

¹⁷ HAMMERSTEIN (2003b) montre même que les conditions nécessaires à la réciprocité (et tout particulièrement des formes de réciprocité du type Donnant Donnant) sont très exigeantes du point de vue des capacités cognitives (mémoire, reconnaissance etc.) si bien qu'il est difficile de la trouver dans le monde animal (il n'en va évidemment pas de même pour le monde humain).

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 ».

g. Le signal coûteux

Un peu dans le même ordre d'idée que l'altruisme réciproque, il existe une théorie qui permet d'expliquer certaines actions 'apparemment altruistes'. Il s'agit de la théorie du signal coûteux. L'idée est qu'un individu peut investir de l'énergie à produire un signal qui lui rapportera un avantage individuel en retour. Par exemple, 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).

Notons toutefois que, tout comme pour la réciprocité, les comportements de type 'signal coûteux' ne peuvent pas prétendre au qualificatif d'« altruisme évolutionniste ».

1. Comment l'altruisme doit-il être pensé par rapport aux êtres humains ?

Les êtres humains ont également leurs kamikazes (à l'exemple des aviateurs japonais durant la seconde guerre mondiale) et leurs héros altruistes comme Winkelried ou Mère Theresa. Outre cela on trouve un bon nombre d'exemples de investissements personnels en faveur du bien de ses proches parents (pensez au phénomène des 'enfants rois') ou de la communauté en général (en Floride par exemple, il est courant de voir des entreprises, institutions, voire même des associations d'étudiants s'organiser volontairement pour entretenir des tronçons d'autoroute). Il est intéressant de se demander si les théories considérées jusqu'à maintenant permettent d'expliquer ces cas, du moins apparents, d'altruisme évolutionniste.

A priori, si on admet que nous sommes un produit de l'évolution, tout porte à penser que notre capacité à agir de manière altruiste est en partie le résultat de la pression de la sélection naturelle. Beaucoup de théoriciens de l'évolution pensent que c'est au temps de l'émergence de l'homo sapiens (paléolithique supérieur) que nos comportements altruistes sont apparus et se sont stabilisés. Les études en 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 2003). 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 la sélection de parentèle et de formes évoluées de réciprocité ou de signal coûteux étaient remplies. Dès lors, se pose la question de savoir si ces théories sont, à elles seules et telles que nous les avons considérées jusqu'à maintenant, 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, basé sur la transmission de comportements génétiquement déterminés. Certes, mais la sélection de parentèle fournit une explication bien trop grossière pour rendre compte du comportement humain. L'enfant roi par exemple apparaît dans un contexte culturel bien précis et on ne peut pas dire que le fait de gâter à l'excès sa progéniture est un simple effet phénotypique de notre code génétique.

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, à transmettre et assimiler des pratiques, des règles de comportements et des connaissances. De manière plus générale, elle doit également rendre compte de notre faculté d'apprentissage et de raisonnement sur des choses complexes, dont il semblerait que nous sommes les uniques bénéficiaires.

En bref, schématiquement, nos actions résultent de trois facteurs. Elles 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).¹⁸ Ainsi, pour pouvoir fournir une explication évolutionniste d'un comportement propre aux êtres humains, il faut utiliser 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, même si elle nous permet d'éclairer (du moins en partie) des phénomènes comme celui de l'enfant roi, à première vue, elle reste muette face à d'autres exemples d'altruisme évolutionniste humain comme les actes de bravoure de Winkelried et des kamikazes japonais, l'engagement humanitaire de Mère Theresa ou les sacrifices que beaucoup de personnes sont disposées à faire au profit de grandes collectivités. Nous verrons plus loin que les psychologues évolutionnistes font précisément appel à la sélection de parentèle pour expliquer l'altruisme en faveur d'individus non parents. 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 à l'adresse 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' qui rapporte un avantage à long terme 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 réitérés 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 Donnan Donnan 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. Il faut donc se tourner vers des

¹⁸ Cette catégorisation est due aux anthropologues évolutionnistes Robert BOYD et Peter RICHERSON (1985).

modèles plus complexes, capables de simuler la richesse des interactions humaines. C'est l'objet de la section suivante.

a. L'anthropologie évolutionniste et l'économie empirique au service de la réciprocité

Entrons dans l'ère de la seconde génération de la théorie des jeux avec des modèles développés en anthropologie évolutionniste et en économie empirique. 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 évolutionnistes humains. Leur programme se développe sur trois fronts : l'expérimentation et l'observation empirique, la modélisation sur ordinateur, et l'explication théorique en termes évolutionnistes.

Dans le domaine empirique, des situations sociales sont représentées sous formes 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 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 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 !¹⁹

Les expériences empiriques (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

¹⁹ Au fond il s'agit d'un dilemme assez proche de celui du prisonnier à la différence qu'il implique plus de deux participants.

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.²⁰

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 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 est que, d'une part, plus de 50% des *trustees* donnent en retour, et que, d'autre part, leur contribution est proportionnelle à l'investissement du *truster*. Plus celui-ci est généreux, plus le *trustee* le récompense (FEHR & FISHBACHER 2003).

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. 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 réitéré, 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). Les nouveaux modèles sont également plus complexes ; ils ne modélisent pas uniquement des interactions dyadiques mais souvent des interactions triadiques ou n-adiques ; parfois, ils modélisent même des situations dans lesquelles les individus peuvent

²⁰ Toutefois, si l'on procède à un jeu du bien commun réitéré 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 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 caustique 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).

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 (dans ce cas, un individu peut savoir, même s'il ne l'a jamais rencontré auparavant, si son partenaire d'interaction a agit précédemment de manière coopérative ou non).

Il serait trop long de présenter les détails de ces théories. L'important est de noter qu'elles 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 au prix de certaines prises de risque. La théorie de la « réciprocité indirecte » par exemple, indique qu'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 (PANCHANATHAN & BOYD 2004; FEHR 2004). Au fond, la réciprocité indirecte est une forme de signal coûteux.

Mais alors, il n'est plus question d'altruisme évolutionniste à proprement parler ! Or si l'on pense que les êtres humains sont réellement capables d'agir de manière altruiste en faveur d'individus non parents, alors il nous faut une autre théorie susceptible d'en expliquer l'évolution. Dans les prochaines sections, je vais présenter deux tentatives qui vont dans ce sens.

b. La punition altruiste

Dans le cadre de leurs recherches, les anthropologues évolutionnistes et les économistes empiriques ont mis en évidence la relative stabilité d'une stratégie altruiste évolutionniste typique des êtres hautement sociaux : il s'agit d'un comportement justicier, qui consiste à punir les opportunistes (BOYD *et al.* 2003 ; FEHR & FISCHBACHER 2003).

La « punition altruiste » (on parle aussi de « réciprocité forte ») est l'action de punir les opportunistes, ceux qui profitent de l'aide des coopérateurs sans jamais rendre de service en retour. Elle est 'altruiste', d'une part parce qu'elle est liée à 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.²¹ De plus, la punition altruiste est liée à un profit

²¹ 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

pour autrui ; en effet, des tests empiriques ont montré qu'après une punition, la tendance à la coopération augmente ; en d'autres termes, les opportunistes punis ont tendance à revenir dans le 'droit chemin'. Ainsi, la punition profite aux individus avec lesquels l'opportuniste puni interagira dans le futur. 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.²²

Des expériences empiriques ont montré que les gens utilisent ce type de comportement dans des situations d'interaction sociale. Ernst FEHR et collègues (FEHR & GÄCHTER 2002 ; FEHR & FISCHBACHER 2004) 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 la possibilité de punir. Les résultats empiriques montrent que s'ils en ont la possibilité, beaucoup de sujets sont prêts à punir les opportunistes à leurs propres frais et sans attente de bénéfices en retour.

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 que les altruistes, comment expliquer que les comportements altruistes punitifs aient pu être sélectionnés au fil de l'évolution ? C'est ici qu'intervient l'explication théorique évolutionniste. Il semblerait qu'il y ait trois facteurs complémentaires qui permettent d'expliquer la stabilisation des comportements altruistes punitifs :

Le premier facteur est lié à la moralisation des comportements altruistes punitifs. Par le biais des normes sociales on peut associer une valeur morale aux comportements punitifs ; cette valorisation morale de la punition altruiste s'accompagne d'une *obligation* de punir, valable pour tous les membres du groupe. Et si l'exécution de la punition devient un devoir moral, alors non seulement les opportunistes seront punis mais aussi les individus qui ne punissent pas (même si par ailleurs ce sont des coopérateurs) ; il y a donc punition des non-punisseurs. Un bon nombre d'expériences théoriques et empiriques (BOYD & RICHERSON 1992 ; FEHR & FISCHBACHER 2004) ont montré que le mécanisme de la punition des non-

aucun coût supplémentaire (le punisseur refuse simplement de coopérer au coup suivant); au contraire, il permet d'éviter une perte (se faire exploiter au coup suivant).

²² A ce propos, on remarque que la punition altruiste permet de transformer des caractéristiques hautement altruistes (celles qui induisent des comportements coopératifs dans un monde d'égoïstes) en des caractéristiques avantageuses du point de vue individuel (si la punition est efficace, il vaut plus la peine de coopérer que de faire défection). Ainsi, avec la punition, ce sont moins les comportements coopératifs qui peuvent prétendre à l'altruisme évolutionniste que les comportements coopératifs-punitifs (puisque le fait de punir engendre un coût).

punisseurs renforce à la fois le comportement coopératif et le comportement punitif ; en d'autres termes, il permet de faire monter le taux moyen de coopération dans le groupe.

Le second 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), alors 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 même nul). En comparaison des individus qui ne punissent pas (les non-altruistes), les punisseurs ne seront donc que légèrement désavantagés. 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.

Il faut noter ici que les deux facteurs susmentionnés ne semblent pas suffisants, à eux seuls, pour assurer l'évolution des comportements altruistes punitifs (en particulier le second).²³ C'est ici que les auteurs (BOYD *et al.* 2003) reviennent aux vertus de l'effet de sélection de groupe. A la différence des modèles que nous avons vu précédemment, 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 :

On part du principe que les êtres humains forment des groupes relativement homogènes, composés d'individus qui adhèrent plus ou moins aux mêmes normes sociales. D'autre part, on présuppose que les êtres humains possèdent une 'tendance au conformisme', c'est-à-dire qu'ils adoptent assez facilement les normes qui ont beaucoup de succès dans leur société (BOYD & RICHERSON 1985). Ainsi, au fil des générations, on observera à l'intérieur de chaque groupe, une tendance à l'uniformisation des normes acceptées (HENRICH & BOYD 2001). 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, il faut que deux conditions soient réunies : premièrement, il faut une variation entre les normes sociales prônées dans les

²³ 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 ; 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.

Quant au facteur de la punition des non-punisseurs, on se trouve devant la difficulté d'une régression à l'infini : il vaut mieux être simple punisseur d'opportunistes plutôt que punisseur à la fois d'opportunistes et de coopérateurs non-punisseurs, etc.

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. 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 coopératifs-punitifs se répandront dans l'ensemble de la population.

En résumé, voici ce qui se passe pour le comportement altruiste punitif. Ce comportement, par définition, 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 au niveau du groupe, ce qui confère à ce dernier un avantage sélectif non négligeable. Ainsi, si au niveau de la sélection individuelle le désavantage engendré par un comportement altruiste punitif n'est pas trop grand, un petit effet de sélection culturelle de groupe suffit à faire pencher la balance à l'avantage du comportement punitif-altruiste.²⁴

Il va sans dire que ce genre de théories, de par leur complexité et leur aspect spéculatif, sont l'objet de nombreuses critiques. Aujourd'hui, le débat est ouvert entre leurs partisans et leurs détracteurs (pour une analyse critique dans ce volume, voir BAUMARD, DE SOUSA, FAUCHER, PERRIN).

²⁴ Notons qu'au niveau culturel, la théorie de la sélection 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).

c. La sélection de parentèle utilisée par la psychologie évolutionniste

La punition altruiste est une chose, mais elle ne saurait couvrir tous les cas d'altruisme que l'on rencontre chez nos pairs. Mère Theresa et Winkelried ne sont certainement pas des 'punisseurs altruistes'. Nous sommes donc toujours en mal d'une explication complète de l'altruisme évolutionniste 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 l'origine des mécanismes proximaux²⁵ 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 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.²⁶ 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.²⁷

A nouveau, cette théorie ne fait pas l'unanimité.²⁸ D'une part, elle repose sur le présumé qu'au pléistocène, les groupes étaient majoritairement constitués de proches parents (ce qu'il est cependant très raisonnable de penser) ; mais surtout, il est difficile d'imaginer que des mécanismes liés à la sélection de parentèle soient dissociés de la capacité à distinguer les proches parents alors que les hommes du pléistocène possédaient très probablement cette capacité.

²⁵ Les mécanismes proximaux sont, à l'échelle des individus, les causes directes des comportements.

²⁶ Ici, il faut distinguer entre le fait de posséder une capacité (en l'occurrence celle à 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).

²⁷ 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 avec un taux d'immigration supérieur, si bien que ces mécanismes et le comportement altruiste qu'ils induisent ont probablement perdu leur vertu adaptative.

²⁸ Pour une analyse critique de la psychologie évolutionniste, voir DEONNA, DE SOUSA, FAUCHER (ce volume).

En fin de compte force est d'admettre que le problème de l'altruisme évolutionniste est loin d'être résolu. D'importantes avancées théoriques ont déjà été faites mais d'autres se font encore attendre.

2. *Désillusion et ouverture*

Nous avons parcouru une série d'explications relatives à l'évolution de l'altruisme évolutionniste. Mais il faut noter que ces approches nous fournissent peu d'indications sur l'altruisme tel qu'il est conçu dans le domaine du sens commun tout comme en psychologie ou en philosophie. En effet, si on n'est pas biologiste ou théoricien des jeux, on considère généralement qu'un acte est altruiste s'il a été causé par des *motifs* dirigés vers le bien d'autrui *et* sans prise en considération d'un quelconque avantage personnel futur. En revanche, nous avons vu que pour un biologiste, un comportement est altruiste s'il a pour *effet* d'augmenter la valeur de survie et de reproduction (*fitness*) d'autrui aux dépens de sa propre valeur de survie et de reproduction (sans prise en compte des motifs). En réalité, ces deux notions sont logiquement distinctes ;²⁹ toute action altruiste évolutionniste ne mène pas forcément à l'altruisme psychologique³⁰ et inversement³¹. Il est très important de garder cette distinction à l'esprit lorsque l'on tente d'utiliser les données biologiques pour expliquer la pensée et le comportement moral.

Les pessimistes diront qu'en fin de compte, toutes les théories mentionnées dans ce chapitre ne nous disent rien sur l'altruisme psychologique pas plus que sur la morale. Mais c'est aller un peu vite. Ce qu'elles nous apportent, c'est une meilleure compréhension de la manière dont fonctionnent les interactions sociales. Elles nous fournissent également un excellent cadre conceptuel dans lequel il est possible de penser l'évolution de caractéristiques intimement liées à la morale ; les émotions morales (voir DE SOUSA, FAUCHER ce volume), les normes sociales (voir BAUMARD ce volume), voire même l'altruisme psychologique. Concernant ce dernier, notons que TOOBY et COSMIDES, en introduisant des mécanismes proximaux comme les émotions empathiques (qui causeraient l'altruisme évolutionniste), mettent en jeu quelque chose de très ressemblant à l'altruisme psychologique.

²⁹ Cette distinction a été très clairement mise en évidence par WILSON et SOBER (2003).

³⁰ C'est par exemple le cas des abeilles kamikazes qui ne conçoivent pas un désir de sauver leurs consœurs avant de piquer l'ours qui s'approche du nid.

³¹ Par exemple si l'on offre généreusement son verre à une personne assoiffée sans savoir que le liquide contient un poison mortel qui nous était destiné

Pour le plaisir, terminons avec une dernière spéculation : Imaginons qu'à un moment de l'évolution 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 (en agissant instinctivement de manière aimable, on s'épargne les coûts de la sanction et des délibérations machiavéliques). Ainsi l'altruisme psychologique aurait pu évoluer parce qu'il est un moyen heuristique efficace pour naviguer dans des environnements sociaux complexes composés de normes renforcées par la punition...

L'histoire ne s'achève évidemment pas ici : objections et spéculations futures sont au programme.*

* Tous mes remerciements à Nicolas PERRIN, mon 'mentor évolutionniste', pour la deuxième partie de cette introduction. Merci surtout à Jacques DUBOCHET pour son grand engagement et son soutien paternel tout au long de ce projet.

A. Bibliographie

AXELROD, Robert, 1996 (1984), *Comment réussir dans un monde d'égoïstes ?*, trad. de l'angl. par M. GARENE, Paris : Odile Jacob.

BATSON, Daniel, SHAW, Laura, 1991, « Encouraging Words Concerning the Evidence for Altruism », *Psychological Inquiry*, 2/2, pp. 159-168.

BOEHM, Christopher, 2002 (2000), « Conflict and the Evolution of Social Control », in L. KATZ (éd.), *Evolutionary Origins of Morality*, *Journal of Consciousness Studies*, 7/1, pp. 79-101.

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*, 100/6, pp. 3531-3535.

BOYD, Robert, RICHERSON, Peter, 1985, *Culture and the evolutionary process*, Chicago: University of Chicago Press.

BOYD, Robert, RICHERSON, Peter, 1992, « Punishment Allows the Evolution of

Cooperation (or anything else) in Sizable Groups », *Ethology and Sociobiology*, 13, pp. 171-195.

BOWLES, Samuel, BOYD, Robert, GINTIS, Herbert, RICHERSON, Peter, 2003, « The Evolution of Altruistic Punishment », *Proceedings of the National Academy of Sciences of the United States of America*, 100/6, pp. 3531-3535.

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.

COSMIDES, Leda, TOOBY, John, 2000. « Evolutionary Psychology and the Emotions », in J. HAVILAND-JONES, M. LEWIS (éd.), *Handbook of Emotions*, New York, London : Guilford, pp. 91-115.

DARWIN, Charles, 1859, *L'origine des espèces*, trad. de l'angl. par E. BARBIER, édité 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.

DAWKINS, Richard, 1996 (1976), *Le Gène Egoïste*, trad. de l'angl. par L. OVION, Paris : Odile Jacob.

DENNETT, Daniel, 2000 (1995), *Darwin est-il dangereux?*, trad. de l'angl. par P. ENGEL, Paris: Odile Jacob.

DE VRIES, Hugo, 1900, « Sur la loi de disjonction des Hybrides », *Comptes Rendus de l'Académie des Sciences*, 130, pp. 845-847.

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.

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, p. 785-791.

FEHR, Ernst, FISCHBACHER, Urs, 2004, « Third-party Punishment and Social Norms », *Evolution and Human Behavior*, 25, pp. 63-87.

FEHR, Ernst, GÄCHTER, Simon, 2002, « Altruistic Punishment in Humans », *Nature*, 415, pp. 137-140.

FISCHBACHER, Urs, GÄCHTER, Simon, FEHR, Ernst, 2001, « Are People Conditionally Cooperative? Evidence From a Public Goods Experiment », *Economics Letters*, 71, pp. 397-404.

GIBBARD, Allan, 1996 (1990), *Sagesse des Choix, justesse des Sentiments : une Théorie du Jugement normatif*, trad. de l'angl. par S. LAUGIER, Paris : PUF.

GILDENHUYS, Peter, 2003, « The Evolution of Altruism: The Sober/Wilson Model », *Philosophy of Science*, 70, pp. 27-48.

HAMILTON, William, 1964, « The Genetical Evolution of Social Behaviour (I and II) », *Journal of Theoretical Biology*, 7, pp. 1-52.

HAMMERSTEIN, Peter, 2003, "Why Is Reciprocity So Rare in Social Animals?", in P. HAMMERSTEIN (éd.), *Genetic and Cultural Evolution of Cooperation*, Cambridge : MIT Press, pp. 83-94.

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.

KORSGAARD, Christine, 1983, « Two Distinctions in Goodness », *The Philosophical Review*, 92/2, pp. 169-195.

KREBS, John, DAVIES, Nicholas, 1993 (1981), *An Introduction to Behavioural Ecology*, Oxford, London: Blackwell.

MAYNARD SMITH, John, 1982, *Evolution and the Theory of Games*, Cambridge: Cambridge University Press.

MAYNARD SMITH, John, 1976, « Group Selection », *The Quarterly Review of Biology*, 51/2, pp. 277-283.

NICHOLS, Shaun, 2004, *Sentimental Rules: On the Natural Foundations of Moral Judgment*, Oxford : Oxford University Press.

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.

OSTROM, Elinor, 1998, « A Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association, 1997 », in *American Political Science Review*, 92/1, pp. 1-22.

PANCHANATHAN, Karthik, BOYD, Robert, 2004, « Indirect Reciprocity Can Stabilize Cooperation Without the Second-Order Free Rider Problem », *Nature*, 432, pp. 499-502.

RACHELS, James, 1990, *Created from Animals; The moral Implications of Darwinism*, Oxford, New York : Oxford University Press.

RICHARDS, Robert, 1993, « Birth, Death, and Resurrection of Evolutionary Ethics », in H. & D. NITECKI (éd.), *Evolutionary Ethics*, Albany: SUNY Press, pp. 113-131.

ROBERTS, Gilbert, SHERRATT, Thomas, 1998, « Development of Cooperative Relationships Through Increasing Investment », *Nature*, 394, pp. 175-179.

ROTTSCHAEFER, William, 1998 (1997), *The Biology and Psychology of Moral Agency*, Cambridge : Cambridge University Press.

SAMUELS, Richard, STICH, Stephen, FAUCHER, Luc, 2004, « Reason and Rationality », in I. NIINILUOTO *et al.* (éd.), *Handbook of Epistemology*, Dordrecht : Kluwer Academic Publishers, pp. 279-300.

STERELNY, Kim, 2003, *Thought in a Hostile World: The Evolution of Human Cognition*, Oxford: Blackwell Publishing

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.

TRIVERS, Robert, 1971, « The Evolution of Reciprocal Altruism », *The Quarterly Review of Biology*, 46, 1, pp. 35-57.

WILKINSON, Gerald, 1984, « Reciprocal Food Sharing in the Vampire Bat », *Nature*, 308, pp. 181-184.

WILSON, David S., SOBER, Elliott, 2003 (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior*, London: Harvard University Press.

WILSON, David S., 1975, « A General Theory of Group Selection », *Proceedings of the National Academy of Sciences*, 72, pp. 143-146.

ZAHAVI, Amotz, 1977, « Reliability in Communication Systems and the Evolution of Altruism », in B. STONEHOUSE *et al.* (éd.), *Evolutionary Ecology*, London: Macmillan Press, pp. 253-259.