

Éthique et intuitions*

Peter Singer

(Princeton University, Charles Sturt University)

Traduction et présentation par

Maxime Gaborit[♦]

Présentation

Précisons-le d'entrée de jeu, la démarche adoptée par Peter Singer dans « Éthique et intuitions » est fondamentalement méthodologique et inachevée. De son propre aveu, l'article n'est qu'une esquisse²⁰⁸, un travail qui demande à être complété et précisé. Son intérêt essentiel est de poser la question de l'importance à accorder à nos intuitions ordinaires dans le cadre de théories morales normatives. Ces intuitions morales correspondent aux jugements évaluatifs immédiats et évidents par eux-mêmes que nous formons spontanément lorsque nous sommes confrontés à une situation particulière. Il existe en philosophie morale plusieurs positions quant au traitement de ces intuitions et quant au rôle que celles-ci doivent jouer dans la formation de règles de conduite. L'utilitarisme dont Singer est l'un des représentants se distingue notamment par son caractère profondément contre-intuitif, la proposition de règles et de normes allant à l'encontre du sens commun est alors souvent l'occasion de réactions énergiques²⁰⁹. Dans le texte ici présenté, Singer est ouvertement critique à l'égard des intuitions ordinaires. Pour autant, on chercherait en vain dans « Éthique et intuitions » des conclusions normatives et définitives spécifiquement utilitaristes. En effet, dans la mesure où le propos de Singer consiste principalement à

* Article initialement publié in *The Journal of Ethics*, 9 (3-4), Dordrecht, Springer Netherlands, 2005, p. 331-352. Nous tenons à remercier Peter Singer d'avoir aimablement autorisé la publication de cette traduction, ainsi que Patrick Lang pour sa relecture et ses précieux conseils (N.d.T.).

♦ Maxime Gaborit suit actuellement des études de philosophie et de lettres modernes à l'Université de Nantes. Parallèlement à ses recherches en philosophie morale et politique, il s'intéresse à l'anti-humanisme théorique, de ses déclinaisons au sein du champ métaphysique jusqu'à ses manifestations dans l'épistémologie des sciences sociales.

²⁰⁸ Voir note 41.

²⁰⁹ Les polémiques à ce sujet ne manquent pas, on peut par exemple penser à Elizabeth Anscombe et son article *La philosophie morale moderne*, trad. de l'anglais par Geneviève Ginvert et Patrick Ducray, *Klesis – Revue Philosophique en Ligne* n°9, 2008, 9-31, dans lequel elle qualifie un homme capable de concevoir qu'il est possible de faire condamner un innocent pour sauver un plus grand nombre de personnes « d'esprit corrompu ».

souligner la nécessité d'une réflexion sur la nature des intuitions et sur leur articulation à une théorie, les distinctions effectuées et les critiques méthodologiques du philosophe australien relèvent de la métaéthique et s'appliquent sans distinction à toute théorie morale.

Cette remise en question des intuitions ordinaires s'inscrit dans le sillage de recherches antérieures, notamment présentées dans *The Expanding circle*²¹⁰ (publié en 1981, réédité en 2011), ouvrage dont « Éthique et intuitions » reprend plusieurs passages. Les deux textes ont pour point commun d'aborder la morale tout en l'associant aux neurosciences et aux théories évolutionnistes. Il ne s'agit pourtant en aucun cas de « biologiser » la morale, comme le ferait par exemple un darwinisme social ; comme Singer prend bien soin de le préciser, il est possible d'adopter une approche évolutive du comportement humain et de s'intéresser à l'éthique sans pour autant défendre une « éthique évolutive », qui passerait abusivement de la formulation de faits scientifiques à l'énonciation de règles d'action. Les découvertes scientifiques, comme Singer n'a de cesse de le préciser, doivent se voir refuser toute portée normative *directe* ; l'établissement scientifique de faits et de relations ne peut suffire à produire des directives éthiques²¹¹ ou à justifier une politique. Tout en appelant à intégrer notre connaissance scientifique à la réflexion morale, Singer évite soigneusement les deux positions extrêmes que seraient d'un côté l'affirmation d'une incompatibilité totale entre le domaine moral et le domaine scientifique, et de l'autre la subordination de la morale aux avancées de la biologie, des neurosciences et de la science en général.

Pour Singer, nous sommes aujourd'hui en possession d'un avantage considérable sur les philosophes qui nous ont précédé, cet avantage consistant en « notre compréhension de l'évolution et son application à l'éthique ». Nous l'avons vu auparavant, la morale a une spécificité qui lui est propre ; la science, quant à elle, offre des outils non négligeables pour comprendre l'éthique et nos propres pratiques. La question sera alors de savoir si cette connaissance peut nous aider d'une façon ou d'une autre à trouver des règles d'action. Afin d'éclaircir ce problème, Singer procède à une généalogie de la moralité, en s'intéressant particulièrement à la nature de nos intuitions morales, à leur origine ainsi qu'à leur traitement philosophique traditionnel. Cette remontée à l'origine le conduit à s'intéresser tout particulièrement à la forme de justice manifeste dans les comportements réciproques observables chez plusieurs espèces animales, et

²¹⁰ P. Singer, *The Expanding Circle*, Princeton, Princeton University Press, 1981.

²¹¹ Les travaux de Hume sur la question ont démontré l'impossibilité de tout passage d'une description de ce qui *est* à l'affirmation d'un *doit être* : voir à ce sujet D. Hume, *Traité de la nature humaine*, Livre III, Paris, Flammarion, 1993.

à proposer une explication évolutionniste de nos comportements et de nos jugements moraux. Il définira alors ceux-ci comme faisant partie intégrante de notre nature biologique, comme le produit d'un long processus de sélection naturelle. En ce sens, « les règles les plus détaillées de la justice, typiques des sociétés humaines pourvues du langage, sont des raffinements du sens instinctif de réciprocité, et peuvent donc être considérées comme artificielles. » Cette nature artificielle n'est pas pour autant un pur arbitraire : en s'appuyant sur Hume, Singer montre qu'éclairer la nature historique et conventionnelle des intuitions morales et des théories qui en découlent ne revient pas pour autant à disqualifier ces dernières. Pour les accepter, il nous faudra toutefois analyser ce qui, en elles, relève de l'évident et du nécessaire. Par la mise en évidence de l'héritage évolutif, il ne s'agit donc pas de tomber dans un relativisme ou d'appeler à dresser une idiosyncrasie des théoriciens moraux, mais plutôt de dénoncer la prétention universelle de certaines théories lorsque celles-ci s'évaluent sur le critère de la correspondance aux intuitions ordinaires, alors prises comme des données absolument déterminantes. Par conséquent, « les progrès en ce qui concerne notre compréhension de l'éthique n'impliquent pas eux-mêmes directement de conclusions normatives, mais ils sapent certaines conceptions quant à la manière de faire de l'éthique, qui ont elles-mêmes des conclusions normatives ». La portée des avancées scientifiques est donc *indirecte*.

« Plus évolué ne signifie pas meilleur ». Le simple fait d'avoir une intuition spécifique, héritée de millénaires d'évolution, ne suffit pas à rendre cette intuition suffisante pour fonder une morale rationnelle. En s'intéressant aux travaux du psychologue Jonathan Haidt, Singer s'attache à montrer que « lorsque se présente un raisonnement conscient plus délibéré, celui-ci a tendance à venir après la réaction intuitive et à être une rationalisation de cette réaction plutôt que la base du jugement moral ». Ce constat n'est pas sans rappeler une des *Pensées* de Pascal :

M. de Roannez disait : « Les raisons me viennent après, mais d'abord la chose m'agrée ou me choque, sans en savoir la raison, et cependant cela me choque par cette raison que je ne découvre qu'ensuite. »

Mais je crois, non pas que cela choquait par ces raisons qu'on trouve après, mais qu'on ne trouve ces raisons que parce que cela choque²¹².

Pascal met ainsi en relief un certain besoin, une certaine tentation de

²¹² B. Pascal, *Pensées*, Brunschvicg 276 / Lafuma 983.

rationalisation de nos intuitions : lorsque nous sommes choqués, il n'est en l'occurrence pas question de *découvrir a posteriori* les raisons du choc éprouvé intuitivement, mais de le *rationaliser*. Singer rejoint cette conclusion : « Dans la vie quotidienne, comme le signale Haidt, notre faculté de raisonner est susceptible de n'être rien de plus qu'une rationalisation de nos réactions intuitives – comme il le formule, le chien émotionnel fait remuer sa queue rationnelle ». Une philosophie morale, par l'union de sa capacité d'abstraction et de son attachement aux intuitions ordinaires, est alors susceptible de devenir un moyen de justification de l'orthodoxie, un outil de légitimation de l'opinion et de réactions non rationnelles.

Sur la base de ces seules considérations, toute tentative de fonder rationnellement une morale à partir de nos intuitions semble alors vouée à l'échec. Mais à l'aide des travaux en neuroscience de Joshua Greene, Singer dégage un élément capable de nous faire sortir de cette aporie : parmi nos intuitions, certaines se distinguent des autres en ce qu'elles présentent un caractère intrinsèquement rationnel. Penser que « la mort d'une personne est une tragédie moindre que la mort de cinq personnes » ou que « le fait qu'une personne soit tuée est une mauvaise chose » peut se réaliser de manière tout à fait intuitive, sans pouvoir pour autant être explicable par notre passé évolutif. Ces intuitions sont le résultat d'une réflexion plus poussée qu'une simple réaction émotive, et permettent alors d'espérer pouvoir fonder rationnellement une théorie morale. Avec « Éthique et intuitions », Singer déplace donc l'enjeu méthodologique en philosophie morale : l'effort n'est plus à concentrer sur l'adéquation de la théorie aux intuitions ordinaires, mais sur la possibilité d'une morale fondée sur la raison pure. Le texte se termine alors sur un constat sans appel, où trois perspectives s'offrent à nous : le scepticisme moral, le nihilisme moral, ou « la tâche ambitieuse de séparer les jugements moraux que nous devons à notre histoire évolutive et culturelle de ceux qui ont une base rationnelle ».

RÉSUMÉ. Pendant des millénaires, les philosophes ont spéculé sur les origines de l'éthique. La recherche récente en psychologie évolutionniste et en neurosciences a éclairci cette question. Mais cette recherche a également une portée normative. Une manière classique d'argumenter contre une théorie éthique normative est de montrer que dans certaines circonstances la théorie mène à des jugements contraires à nos intuitions morales ordinaires. Si, cependant, ces intuitions morales sont le résidu biologique de notre histoire évolutive, la raison selon laquelle nous devrions les regarder comme ayant une quelconque force normative n'est pas claire. La recherche en neurosciences devrait par conséquent nous mener à reconsidérer le rôle des intuitions en éthique normative.

MOTS-CLÉS : David Hume, dilemme du tramway, Emmanuel Kant, éthique, Henry Sidgwick, imagerie cérébrale, intuitions, James Rachels, John Rawls, Jonathan Haidt, Joshua D. Greene, neuroscience, psychologie évolutionniste, utilitarisme.

I. Introduction

Dans l'un de ses nombreux excellents essais, Jim Rachels critiquait les philosophes « prompts à dégainer ». Comme il le disait :

Le téléphone sonne, et un journaliste débite à toute allure quelques “événements” à propos de quelque chose que quelqu'un est supposé avoir fait. Des problèmes éthiques sont impliqués – quelque chose d'alarmant est dit avoir eu lieu – et on demande alors à l'« éthicien » un commentaire à inclure dans l'article du lendemain, qui peut être le premier compte-rendu que le public aura vu au sujet des événements en question²¹³.

Dans ces conditions, remarquait Rachels, le reporter veut une citation concise, disant de préférence que les événements décrits sont graves. Le philosophe effectue un jugement en deux coups de cuillère à pot, et le résultat est quelque chose qui reflète non pas une « analyse prudente » mais une « sagesse acceptée » (*accepted wisdom*). Les philosophes deviennent les défenseurs les plus sophistiqués de l'orthodoxie, supposant que le consensus social existant doit être conforme à la morale, et exprimant clairement sa

²¹³ J. Rachels, « When Philosophers Shoot from the Hip », in *Bioethics: An Anthology*, Oxford (éd. Helga Kuhse et Peter Singer), Blackwell Publishers, 1999, p. 573.

« justification » théorique. Rachels soutenait que les philosophes doivent au contraire « remettre en cause l'orthodoxie dominante, en remettant en question les hypothèses que les gens font sans réfléchir »²¹⁴.

Le propre travail de Rachels en éthique était en accord avec ce précepte. Pour donner un exemple parmi tant d'autres, dans ce qui est probablement son article le plus cité (sur l'euthanasie active et l'euthanasie passive), il s'est exprimé pour critiquer l'intuition ordinaire selon laquelle tuer est pire que laisser mourir. Il a montré que cette distinction est influente en médecine, et est incorporée à une déclaration de l'Association médicale américaine. Il a soutenu ensuite de manière convaincante que ce n'est pas une intuition à laquelle nous devrions nous fier²¹⁵.

Dans les deux articles que j'ai mentionnés, Rachels rejetait l'idée que le rôle des philosophes moraux est de prendre nos intuitions morales ordinaires comme des données, et de chercher à développer la théorie qui correspond le mieux à ces intuitions. Au contraire, maintient-il, nous devrions être prêts à remettre en cause les intuitions qui nous sont d'abord venues à l'esprit lorsqu'on nous a demandé de nous prononcer sur un problème moral. C'est un point de vue que je partage, et sur lequel j'ai écrit en de multiples occasions au fil des années²¹⁶. Dans les pages suivantes, je soutiens que la recherche récente en neurosciences nous donne de nouvelles et puissantes raisons d'adopter une position critique à l'égard des intuitions ordinaires. Mais je commencerai par situer cette recherche dans le contexte de notre longue quête des origines et de la nature de la moralité.

Il y a dans le musée du Louvre une colonne babylonienne noire avec un relief montrant le dieu du soleil Shamash en train de donner le code des lois à Hammurabi. De tels récits mythiques, attribuant une origine divine à la moralité, sont courants. Dans le *Protagoras* de Platon, il y a un récit, de son propre aveu mythique, de la manière dont Zeus prit pitié des humains infortunés, qui, vivant en petit groupes et avec des dents inadaptées, de mauvaises griffes et un manque de rapidité, ne pouvaient égaler les autres bêtes. Pour compenser ces déficiences, Zeus donna aux hommes un sens moral et une aptitude à la loi et à la justice, pour qu'ils puissent vivre dans de plus larges communautés et coopérer les uns avec les autres. La relation biblique de Dieu donnant les Dix Commandements à Moïse sur le Mont Sinaï est, bien sûr, un autre exemple.

Outre ces récits mythiques, pendant au moins 2500 ans et dans

²¹⁴ *Idem*, p. 575.

²¹⁵ J. Rachels, « Active and Passive Euthanasia », in *Bioethics: An Anthology*, Oxford (éd. Helga Kuhse et Peter Singer), Blackwell Publishers, 1999, p. 227–230.

²¹⁶ P. Singer, « Sidgwick and Reflective Equilibrium », in *The Monist*, 58, 1974, p. 490-517.

différentes civilisations, les philosophes ont débattu et écrit sur la nature de l'éthique. Platon lui-même n'était à l'évidence pas satisfait de l'explication qu'il propose dans le *Protagoras*, car il examine dans ses dialogues plusieurs autres possibilités. Dans la seule *République*, nous trouvons la déclaration sceptique de Thrasymaque selon lequel les forts, qui agissent dans leur propre intérêt, imposent ce qui est moral aux faibles, nous trouvons le modèle du contrat social de Glaucon, et nous trouvons la défense de Socrate de la justice selon la loi proto-naturelle, justice comprise comme résultat de l'harmonie des différentes parties de la nature humaine.

Parmi les questions auxquelles ont réfléchi les philosophes, on trouve celles de savoir si l'éthique est objectivement vraie, ou relative à la culture, ou entièrement subjective ; si les êtres humains sont naturellement bons ; et si l'éthique vient de la nature ou de la culture. Les philosophes ont considéré de telles questions comme ayant une importance pratique, autant que théorique. Y répondre, croient-ils, nous permettra de vivre d'une meilleure manière. Beaucoup de ces penseurs étaient d'habiles observateurs de leurs semblables, outre qu'ils étaient parmi les personnes les plus sages de leur époque. Considérez, par exemple, le travail de Mencius, d'Aristote, de Nicolas Machiavel, de Thomas Hobbes et de David Hume. Il y a beaucoup de choses concernant la nature humaine qu'ils ont très bien saisi. Mais aucun d'eux n'avait l'avantage d'une approche scientifique moderne de ces problèmes. Nous avons aujourd'hui cet avantage. C'est pourquoi il serait surprenant que nous ne puissions pas perfectionner ce qu'ils ont écrit.

Dans ce qui suit, je résume une partie de la connaissance nouvelle que nous avons maintenant de l'éthique, connaissance qui n'était disponible pour aucun des grands philosophes que j'ai nommés. Puis j'examinerai quelle importance normative a cette nouvelle connaissance. Que pourrait apporter cette connaissance à notre débat sur la question de savoir comment nous devrions agir – si toutefois elle peut y contribuer en quoi que ce soit ?

II. La théorie évolutionniste et les origines de la moralité

L'avantage particulier le plus important que nous ayons sur les grands philosophes moraux du passé est notre compréhension de l'évolution et son application à l'éthique. Bien que les philosophes que j'ai mentionnés fussent capables de se libérer eux-mêmes du mythe de l'origine divine de la moralité et d'expliquer la moralité en termes naturalistes, ils manquaient d'une compréhension exacte de la façon dont nos normes ont pu apparaître par sélection naturelle avec le gène comme unité fondamentale pour la transmission des caractères héréditaires entre générations. Sans ce savoir, ils

pouvaient observer nos sentiments et attitudes mais ne pouvaient pas les expliquer de manière adéquate. Pour voir ce que la théorie évolutionniste peut apporter même aux plus grands des penseurs pré-Darwiniens ayant spéculé à propos des origines de la moralité, considérez le traitement de la moralité par Hume dans son *Traité de la nature humaine*, célébré à juste titre.

Hume ouvre son examen de la justice en posant la question de savoir si la justice est une vertu naturelle ou artificielle. En examinant cette question, il écrit :

Un homme aime naturellement mieux ses enfants que ses neveux, ses neveux mieux que ses cousins, ses cousins mieux que des étrangers, toutes choses égales par ailleurs. De là viennent nos définitions courantes du devoir qu'il y a à préférer les uns aux autres. Notre sens du devoir suit toujours le cours naturel et ordinaire de nos passions²¹⁷.

Hume s'approche de très près d'une compréhension évolutionniste du sens ordinaire du devoir, mais ne pouvait expliquer comme le peut la théorie évolutionniste moderne pourquoi « le cours naturel et ordinaire de nos passions » prend cette forme particulière. Nous comprenons maintenant que les gènes menant aux formes d'amour que Hume décrit sont plus susceptibles de survivre et de se répandre parmi les mammifères sociaux que les gènes qui ne conduisent pas à des préférences pour des parents typiquement proportionnelles à la proximité de la relation de parenté. Car nous partageons plus de gènes avec nos enfants qu'avec nos cousins, et plus avec nos cousins qu'avec des étrangers.

Nous pouvons maintenant aussi fournir une explication plus profonde de la justesse de l'observation en sens inverse, et plus contestable de Hume, selon laquelle « il n'y a pas dans l'esprit de l'homme une passion telle que l'amour de l'humanité, uniquement comme telle, sans rapport avec certaines qualités personnelles, avec les services qu'on nous a rendus ou avec une relation à nous-mêmes²¹⁸ ». Si regrettable que cela soit, la plupart des êtres humains manque d'un sentiment général de bienveillance envers les étrangers qu'ils croisent dans la rue. En termes évolutionnistes, lorsque l'on considère l'espèce comme un tout, l'unité de sélection est trop large pour que la sélection naturelle ait beaucoup d'impact. En dépit des albums illustrés que nous avons enfants, la vie humaine à ses débuts n'était pas, de

²¹⁷ D. Hume, *Traité de la nature humaine*, Livre III, Partie 2, Section 1, Paris, Flammarion, 1993, p. 82.

²¹⁸ *Idem*, p. 79-80.

manière générale, une lutte pour la survie entre hommes et tigres à dents de sabre. C'était beaucoup plus souvent une lutte pour la survie entre différents êtres humains. Le simple fait que les autres soient membres de notre espèce ne suffit pas à donner à notre souci à leur égard un avantage évolutionniste. À la différence de la sélection d'organismes individuels au sein de l'espèce, qui se poursuit incessamment, la sélection entre espèces différentes arrive trop lentement et trop rarement pour jouer un grand rôle dans l'évolution.

Notez, cependant, les facteurs qu'énumère Hume comme produisant de l'amour à l'égard des autres : les qualités personnelles, les services, et la relation à soi-même. Nous avons déjà traité de la parenté. Les qualités personnelles peuvent produire des sentiments positifs car elles sont susceptibles de nous être bénéfiques, à nous ou à un petit groupe auquel nous appartenons. À la différence de la sélection entre espèces, qui est rare et de peu d'importance dans l'évolution, la sélection au sein de l'espèce, entre de plus petits groupes de reproduction isolés, arrive beaucoup plus fréquemment. Ces plus petits groupes rivalisent entre eux, et, en comparaison avec l'espèce, sont relativement éphémères. Les pressions de sélection faisant contrepoids au niveau de l'individu ou du gène s'appliqueraient toujours, mais moins efficacement. Dans certaines circonstances, il pourrait y avoir des pressions sélectives qui favorisent le sacrifice personnel au profit du groupe. Il y aurait aussi, bien sûr, des pressions en sens inverse favorisant les actions motivées par un intérêt personnel qui ne profitent pas au groupe. Si, toutefois, le groupe développe une culture qui récompense ceux qui risquent leurs propres intérêts afin de profiter au groupe, et qui punit ceux qui ne le font pas, le rapport coût-profit serait incliné de telle manière à faire en sorte que profiter au groupe sera plus susceptible d'être compatible avec le fait de se reproduire à la génération suivante.

La troisième exception que Hume mentionnait était « les services ». Ici encore il aborde un point de théorie évolutionniste récente, qui s'est imbriqué avec la théorie des jeux dans l'exploration de situations telles que le dilemme du prisonnier. Ce travail nous rend capable de donner une réponse plus complète et plus convaincante que Hume ne le pouvait à la question par laquelle il avait commencé son examen de la justice.

Hume se demandait si la justice est une vertu naturelle ou artificielle, et répondit que c'est une vertu artificielle. Par là il voulait dire que « le sens de la justice et de l'injustice ne vient pas de la nature, mais résulte artificiellement, quoique nécessairement, de l'éducation et des conventions humaines ». Il ajoute que bien que les règles de justice soient artificielles, cela ne signifie pas qu'elles soient arbitraires. La justice est, pour Hume,

une invention humaine, bien qu'une invention « évidente et absolument nécessaire »²¹⁹.

Mais la justice n'est pas, du moins à l'origine, une invention humaine. Nous pouvons en trouver des formes dans nos parents non-humains les plus proches. Un singe présentera son dos à un autre singe, qui en enlèvera les parasites ; après quelque temps les rôles seront inversés. Un singe qui ne retourne pas la faveur est susceptible à l'avenir d'être attaqué, ou rejeté. Une telle réciprocité sera payante, en termes évolutionnistes, aussi longtemps que les coûts de l'aide donnée seront inférieurs au bénéfice de l'aide reçue, et aussi longtemps que les animaux n'obtiendront aucun avantage sur le long terme en « trichant » – c'est-à-dire en recevant des faveurs sans les rendre en contrepartie. Il semblerait que la meilleure façon de s'assurer que ceux qui trichent ne prospèrent pas soit pour les animaux d'être capable de reconnaître les tricheurs et de leur refuser pour la prochaine fois l'avantage d'une coopération. Ceci est uniquement possible entre des animaux intelligents vivant en de petits groupes stables sur une longue période. Des preuves confirment cette conclusion : des comportements réciproques ont été observés chez les oiseaux et les mammifères, les cas les plus manifestes survenant parmi les loups, les chiens sauvages, les dauphins, les singes (*monkeys*) et les grands singes (*apes*).

De nombreuses caractéristiques de la moralité humaine pourraient s'être développées à partir de simples pratiques réciproques comme l'extirpation mutuelle de parasites d'endroits difficiles d'accès. Supposez que je veuille qu'on m'enlève les poux qui se trouvent dans mes cheveux et que je sois prêt à ôter en retour les poux des cheveux de quelqu'un d'autre. Je dois cependant choisir mon partenaire avec soin. Si j'aide tout le monde sans distinction, je me retrouverai à épouiller les autres sans que mes propres poux soient enlevés. Pour éviter cela, je dois apprendre à faire la distinction entre ceux qui retournent les faveurs et ceux qui ne le font pas. En effectuant cette distinction, je sépare les réciprocatrices et les non-réciprocatrices²²⁰ et, ce faisant, je développe des notions brutes de justice et de tricherie. Je renforcerai mes liens avec ceux qui me retournent les faveurs, et en résulteront des liens d'amitié et de loyauté, entraînant à leur tour un sens de l'obligation d'aider.

²¹⁹ *Idem*, p. 83.

²²⁰ Le terme de *réciprocatrice* (*reciprocator*) désigne à l'origine un outil servant à débroussailler. Dans ce texte, on prendra bien évidemment le néologisme *réciprocatrice* dans un sens comportemental, le réciprocatrice étant compris comme un être particulier manifestant une tendance à la réciprocité. Par dérivation, le non-réciprocatrice renverra à un être ne manifestant pas cette tendance (N.d.T.).

Ce n'est pas tout. Comme nous le voyons avec les singes, il est probable que les réciprocaturs réagissent d'une manière hostile et colérique à ceux qui ne retournent pas les faveurs. Les réciprocaturs les plus sophistiqués, capables de penser et d'utiliser le langage, peuvent considérer la réciprocité comme bonne et « morale » (*right*), et la tricherie comme mauvaise et « immorale » (*wrong*). De là à conclure que les pires des non-réciprocaturs devraient être chassés de la société, ou être punis d'une autre manière afin qu'ils ne profitent plus abusivement des autres, il n'y a qu'un pas. Ainsi un système de punition et une notion d'isolement constituent l'autre face de l'altruisme réciproque.

Hume n'avait ainsi pas tout à fait tort en disant que la justice est une vertu artificielle, mais il n'avait pas non plus tout à fait raison. La règle fondamentale de la réciprocité, qui inclut la capacité de détecter les tricheurs et le sens de l'indignation requis pour les écarter, est naturelle dans le sens où elle a évolué, où elle fait partie de notre nature biologique, et où elle est quelque chose que nous partageons avec nos parents non-humains les plus proches. Mais les règles les plus détaillées de la justice, typiques des sociétés humaines pourvues du langage, sont des raffinements du sens instinctif de réciprocité, et peuvent donc être considérées comme artificielles.

Notre complexion biologique ne prescrit pas les formes spécifiques que prend notre moralité. Il y a des variations culturelles dans la moralité humaine, comme le savait même Hérodote²²¹. Néanmoins, il semble probable que toutes ces différentes formes sont le développement d'un comportement qui existe chez les animaux sociaux, développement qui est le résultat du processus évolutif habituel de la sélection naturelle. La moralité est un phénomène naturel. Nul mythe n'est requis pour expliquer son existence.

III. Comment les hommes forment des jugements moraux

Une fois posée cette compréhension des origines de la moralité, je me tourne maintenant vers quelques recherches scientifiques récentes qui nous aident à comprendre des décisions et des comportements moraux plus spécifiques. Pour étudier la manière dont les gens parviennent à des jugements moraux, Jonathan Haidt, un psychologue de l'Université de Virginie, a demandé à des gens de réagir à l'histoire suivante :

²²¹ Voir à ce sujet son récit des efforts de Darius, l'empereur perse, pour convaincre les peuples de différentes cultures de changer leurs coutumes quant à la manière de s'occuper des défunts, in Hérodote, *L'Enquête*, Livre III, Chapitre 38, Paris, Gallimard, 1985.

Marc et Julie sont frère et sœur. Ils voyagent ensemble en France lors de leurs vacances d'été. Une nuit, ils séjournent seuls dans une cabane à côté de la plage. Ils décident qu'il pourrait être intéressant et amusant d'essayer de faire l'amour. Ce serait au moins une nouvelle expérience pour chacun d'eux. Julie prenait déjà la pilule, mais Marc utilise en plus un préservatif, pour plus de sûreté. Ils apprécient tous les deux de faire l'amour, mais décident de ne plus le faire à l'avenir. Ils gardent le souvenir de cette nuit comme un secret particulier, qui les fait se sentir encore plus proches l'un de l'autre. Qu'en pensez vous, était-il bien pour eux de faire l'amour ?.

Haidt rapporte que la plupart des gens sont prompts à répondre que ce qu'ont fait Marc et Julie était mal. Ils essaient ensuite de donner des raisons à leur réponse. Il se peut qu'ils mentionnent les dangers de la consanguinité, mais qu'ils se rappellent ensuite que Marc et Julie ont utilisé deux formes de contraception. Ou il se peut qu'ils suggèrent que le frère ou la sœur pourrait être blessé, même s'il est évident d'après l'histoire que cela n'a pas été le cas. Finalement, beaucoup de personnes répondent quelque chose comme : « Je ne sais pas, je ne peux pas l'expliquer, je sais juste que c'est mal²²² ». Bien évidemment, c'est la réaction intuitive qui est responsable du jugement auquel ces gens arrivent, et non pas les raisons qu'ils offrent. En effet, ils restent fidèles à leur jugement immédiat, intuitif, même après avoir soustrait les raisons qu'ils avaient données initialement à ce jugement, et sont incapables d'en trouver de meilleures.

Un exemple pris isolément ne démontrerait pas grand chose, mais Haidt a réuni un ensemble impressionnant de preuves en faveur de la conception selon laquelle les jugements moraux dans divers domaines sont typiquement le résultat de réactions intuitives rapides, presque automatiques. Lorsque se présente un raisonnement conscient plus délibéré, celui-ci a tendance à venir après la réaction intuitive et à être une rationalisation de cette réaction plutôt que la base du jugement moral²²³.

²²² J. Haidt, F. Björklund, S. Murphy, « Moral Dumbfounding: When Intuition Finds No Reason » (Département de Psychologie, Université de Virginie, 2000, manuscrit non publié) ; pour approfondir, se reporter à la discussion qu'en fait J. Haidt dans « The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment », *Psychological Review*, 108, 2001, pp. 814–834. Je suis redevable à Joshua Greene pour avoir attiré mon attention sur ce sujet et sur d'autres matériaux abordés dans cette partie, qui s'appuient sur J. Greene, *The Terrible, Horrible, No Good, Very Bad Truth About Morality, and What to Do About It* (thèse de doctorat, Département de Philosophie, Université de Princeton, 2002), chapitre 3.

²²³ J. Haidt, « The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment », *op. cit.*

Si nous nous tournons vers notre connaissance croissante des parties du cerveau impliquées dans les décisions éthiques, nous trouvons un tableau compatible avec les conclusions que Haidt a tirées d'études du comportement humain. Nous pouvons ici commencer avec l'examen révélateur réalisé par Antonio Damasio, examen d'un cas étant survenu au XIX^e siècle, le cas de Phineas Gage²²⁴. Gage travaillait sur le réseau ferroviaire des États-Unis lorsqu'une explosion fit passer une barre de fer de près d'un mètre de longueur à travers son cerveau. Contre toute attente, Gage survécut et sembla se rétablir complètement, sans détérioration de sa faculté de raisonnement ou de ses capacités linguistiques. Pourtant, il devint peu à peu manifeste que son caractère, auparavant constant et travailleur, avait changé. Il était devenu antisocial, et ne parvenait pas à garder un travail stable comme il le faisait auparavant.

La blessure de Gage avait touché la partie ventromédiane des lobes frontaux. Des patients plus proches de nous historiquement et avec cette zone endommagée montrent la même combinaison de faculté de raisonnement intacte et d'infractions aux usages moraux et sociaux standards. Ces patients semblent être émotionnellement défectueux, dans la mesure où ils ne réagissent pas de manière habituelle à des scènes sanglantes, dans lesquelles les vies d'individus ont été perdues ou menacées. Damasio dit de l'un d'entre eux que sa condition était « de savoir, mais de ne pas ressentir²²⁵ ». Des études en imagerie cérébrale ont montré une corrélation entre le comportement antisocial et une insuffisance, soit dans la taille du cortex préfrontal, soit dans la somme d'activité métabolique de celui-ci²²⁶. Deux patients, dont la région ventromédiane des lobes frontaux avait été endommagée au début de leur vie, montraient des tendances psychopathes bien plus marquées. Ils mentaient, volaient et agissaient de manière violente, et n'éprouvaient aucun remords²²⁷.

Une compréhension plus complète de la manière dont nous formulons des jugements moraux est venue très récemment d'expériences utilisant l'imagerie par résonance magnétique fonctionnelle (ou IRMf), ces expériences ayant été dirigées par Joshua Greene et quelques autres

²²⁴ A. R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*, New York, Grosset&Putnam, 1994, p. 3-9, 34-51.

²²⁵ *Idem*, p. 45.

²²⁶ A. Raine, T. Lencz, S. Bihrlé, L. LaCasse et P. Colletti, « Reduced Prefrontal Gray Matter Volume and Reduced Autonomic Activity in Antisocial Personality Disorder », *Archives of General Psychiatry*, 57, 2000, pp. 119-127 ; A. Raine, M. S. Buchsbaum, J. Stanley, S. Lottenberg, L. Abel et J. Stoddard, « Selective Reductions in Prefrontal Glucose Metabolism in Murderers », *Biological Psychiatry*, 36, 1994, p. 365-373.

²²⁷ S. W. Anderson, A. Bechara, H. Damasio, D. Tranel et A. R. Damasio, « Impairment of Social and Moral Behavior Related to Early Damage in Human Prefrontal Cortex », *Nature Neuroscience*, 2, 1999, p. 1032-1037.

personnes à l'université de Princeton. Greene a conçu les expériences afin de faire la lumière sur la manière dont les gens réagissent à des situations connues en littérature philosophique sous le nom de « dilemmes du tramway²²⁸ ». Dans le dilemme du tramway classiquement posé, vous vous tenez près d'une voie ferrée lorsque vous remarquez qu'un tramway, sans personne à son bord, dévale la voie en se dirigeant vers un groupe de cinq personnes. Elles seront toutes tuées si le tramway continue sur sa lancée. La seule chose que vous pouvez faire pour empêcher ces cinq morts est de tirer un levier qui dirigera le tramway sur les rails voisins, où il tuera une personne seulement. Lorsque l'on demande aux gens ce qu'il faudrait faire dans une telle situation, la plupart estime que l'on devrait dévier le tramway sur la voie voisine, et ainsi sauver quatre vies.

Dans une autre version du problème, le tramway, comme dans la situation précédente, est sur le point de tuer cinq personnes. Cette fois, cependant, vous ne vous tenez pas à côté de la voie, mais sur une passerelle piétonne au dessus de celle-ci. Vous ne pouvez pas dévier le tramway. Vous envisagez de sauter de la passerelle devant le tramway, vous sacrifiant ainsi pour sauver les personnes en danger, mais vous réalisez que vous êtes bien trop léger pour l'arrêter. Cependant, un inconnu très gros se tient juste à côté de vous. Le seul moyen que vous ayez d'empêcher le tramway de tuer cinq personnes est de pousser ce gros inconnu du haut de la passerelle, devant le tramway. Si vous le poussez, il sera tué, mais vous sauverez la vie des cinq autres. Lorsque l'on demande ce qu'il faudrait faire dans de telles circonstances, la plupart des gens estime que l'on ne devrait pas pousser l'inconnu du haut de la passerelle.

Beaucoup de philosophes, dont Judith Jarvis Thomson, saisissent le problème posé par ces deux cas pratiques. Dans les deux cas vous provoquez la mort d'une personne pour en sauver cinq, mais on juge votre action comme bonne dans le dilemme du tramway classique et comme mauvaise dans le cas de la passerelle. Qu'est-ce qui fait la différence entre ces deux situations ? Ces philosophes prennent donc d'emblée les intuitions morales suscitées par les expériences de pensée comme correctes, et cherchent à les justifier. Mais chaque fois qu'un principe justificatif apparemment plausible a été suggéré, d'autres philosophes ont produit des variantes de la paire originale de situations montrant que le principe suggéré

²²⁸ Phillipa Foot semble avoir été la première philosophe à avoir discuté de ces problèmes, in P. Foot, « The Problem of Abortion and the Doctrine of the Double Effect », *Oxford Review*, 5, 1967, p. 5-15 ; réimprimé in J. Rachels (éd.), *Moral Problems: A Collection of Philosophical Essays*, New York, Harper& Row, 1971, p. 28-41. L'article classique sur le sujet reste néanmoins celui de Judith Jarvis Thomson, « Killing, Letting Die, and the Trolley Problem », *The Monist*, 59, 1976, p. 204-217.

ne parvient pas à justifier nos réactions intuitives. Par exemple, certains philosophes ont suggéré que la différence entre le problème standard du tramway et la variante de la passerelle est que dans le dernier cas l'inconnu est utilisé comme un moyen pour sauver les autres. Pousser l'inconnu du haut de la passerelle viole donc l'injonction kantienne de ne pas utiliser une autre personne simplement comme un moyen, tandis que tirer le levier la respecte. Malheureusement pour les partisans de cette explication habile, nous pouvons imaginer un cas où tirer le levier ne fait pas dévaler le tramway sur une voie complètement différente, mais le fait d'effectuer une boucle avant qu'il n'atteigne les cinq personnes menacées. Sur cette boucle est allongé le gros inconnu. Parce qu'il est énorme, son corps arrêtera le tramway, qui le tuera au passage. Dévier le tramway sur cette boucle revient à utiliser l'inconnu comme un moyen pour sauver la vie des cinq autres, mais la plupart des gens considère qu'il serait moralement juste de le faire. Ils jugent ainsi ce cas comme étant plus proche du fait de tirer le levier que du fait de pousser l'étranger du haut de la passerelle.

À la différence de nombreux philosophes qui ont essayé de justifier nos intuitions dans ces situations, Greene était plus intéressé par le fait de comprendre pourquoi nous les avons. Il pensait que les racines des différents jugements que nous formulons sur ces deux situations peuvent résider dans nos différentes réactions émotionnelles à l'idée de causer la mort d'un inconnu en tirant un levier sur une voie ferrée et à l'idée de précipiter à mains nues quelqu'un vers sa mort. Comme le disait Greene :

Parce que les gens ont une réaction émotive robuste, négative, à la violation personnelle proposée dans le cas de la passerelle, ils disent immédiatement qu'elle est mauvaise... Parallèlement, les gens ne parviennent pas à avoir une réaction émotive forte à la violation relativement impersonnelle proposée dans le cas classique du tramway, et reviennent par conséquent au principe moral le plus évident, celui de « minimiser les dégâts », qui les conduit cette fois-ci à dire que l'action est permise dans le cas initial²²⁹.

Greene a utilisé l'imagerie IRMf, qui fournit une image en temps réel de l'activité dans différentes parties du cerveau, pour évaluer cette hypothèse. Il a prédit que les personnes auxquelles on demanderait de formuler un jugement moral sur des violations « personnelles » (comme le fait de

²²⁹ J. Greene, *The Terrible, Horrible, No Good, Very Bad Truth About Morality, and What to Do About It*, p. 178.

pousser l'inconnu de la passerelle) manifesteraient dans des zones du cerveau associées aux émotions une activité intensifiée, par rapport aux personnes auxquelles on aurait demandé de formuler des jugements à propos de violations relativement « impersonnelles » (comme le fait de tirer un levier). Mais Greene a également fait une prévision plus spécifique : celle que la minorité des sujets considérant qu'il serait moral de pousser l'étranger du haut de la passerelle donnerait cette réponse – à moins qu'ils ne soient des psychopathes – en dépit de leurs émotions ; il a prédit que ces mêmes sujets prendraient par conséquent plus de temps pour parvenir à ce jugement que ceux affirmant qu'il serait mauvais de pousser l'inconnu, et que parvenir dans ces conditions à ce jugement prendrait plus de temps que dans une situation qui ne suscite pas une réaction émotionnelle aussi forte.

Les prévisions de Greene ont été confirmées. Lorsque l'on a demandé aux gens de formuler des jugements dans des cas « personnels », les parties de leurs cerveaux associées à l'activité émotionnelle étaient plus actives que lorsque l'on leur a demandé de formuler des jugements dans des cas « impersonnels ». Plus important encore, ceux qui arrivèrent à la conclusion qu'il serait moral d'agir par des moyens qui impliquent une violation personnelle, mais minimisent dans l'ensemble les dommages – par exemple, ceux qui estiment qu'il serait moral de pousser l'inconnu de la passerelle – prirent plus de temps pour formuler leur jugement que ceux estimant qu'il serait mal de le faire²³⁰.

Lorsque Greene a regardé de plus près l'activité cérébrale de ces sujets qui acceptaient des violations personnelles permettant de minimiser la somme des dommages, il a trouvé qu'ils manifestaient une activité plus importante dans les parties du cerveau associées à l'activité cognitive que ceux qui s'opposaient à de telles actions²³¹. Ce sont là des résultats provisoires, reposant sur une quantité limitée de données. Mais supposons qu'ils soient solides, et spéculons sur ce qui pourrait en découler, conformément aux autres informations scientifiques pertinentes quant aux origines de l'éthique exposées précédemment.

²³⁰ J. D. Greene, R. B. Sommerville, L. E. Nystrom, J. M. Darley et J. D. Cohen, « An fMRI Investigation of Emotional Engagement in Moral Judgment », *Science*, 293, 2001, p. 2105-2108. Pour être plus précis : dans les dilemmes moraux personnels, le cortex orbitofrontal médian, le cortex cingulaire postérieur et le gyrus angulaire/sillon temporal supérieur sont actifs. Dans le cas de dilemmes moraux impersonnels, on observe une activité intensifiée dans le cortex préfrontal dorso-latéral et dans le lobe pariétal.

²³¹ J. Greene et J. Haidt, in « How (and Where) Does Moral Judgment Work? », *Trends in Cognitive Sciences*, 6, 2002, p. 517-523, ce à quoi s'ajoutent des conversations personnelles. Pour être plus précis, ceux qui acceptent la violation personnelle montrent plus d'activité dorso-latérale préfrontale antérieure, alors que ceux qui la rejettent ont plus d'activité dans la zone du précuneus.

IV. Implications normatives

Peu après la parution de *L'Origine des espèces*, Darwin écrivit à un ami : « On m'a lancé dans un journal de Manchester un pétard mouillé plutôt bon, montrant que j'avais prouvé que " la raison du plus fort est toujours la meilleure ". »²³². Darwin savait, bien entendu, qu'il n'avait rien prouvé de la sorte. Les darwinistes sociaux ont commis la même erreur lorsqu'ils s'opposèrent à l'intervention de l'État dans l'économie de marché en raison du fait que protéger les pauvres et les faibles serait un obstacle à la sélection naturelle. En supposant que l'on puisse définir le terme « naturel » d'une manière qui rende sensé le fait de dire que protéger les pauvres et les faibles est un obstacle à la sélection naturelle, nous aurions encore besoin d'un argument éthique pour dire qu'il est immoral de le faire. Le sens de l'évolution ne suit pas ni n'a une quelconque connexion nécessaire avec les voies du progrès moral. « Plus évolué » ne signifie pas « meilleur ». Si souvent qu'on ait signalé la fausseté d'une lecture morale du sens de l'évolution, des gens la reproduisent toujours, et il n'est pas difficile de trouver au sein de la théorie de l'évolution des auteurs contemporains – par ailleurs excellents – qui continuent de faire cette erreur. Cela n'en reste pas moins une erreur²³³. Même si j'ai soutenu que la théorie évolutive explique une grande part de la moralité ordinaire, notamment le rôle central des devoirs envers nos proches et des devoirs relatifs à la réciprocité, je ne soutiens pas que cette théorie évolutive justifie ces éléments de la moralité ordinaire. Je suis partisan d'une approche évolutive du comportement humain et je m'intéresse à l'éthique, mais je ne suis pas défenseur d'une « éthique évolutive ».

L'impossibilité de dégager des conclusions éthiques de faits d'évolution ne signifie pas que les récents progrès dans notre compréhension scientifique de l'éthique n'ont pas du tout de portée normative. Ces progrès sont hautement significatifs pour l'éthique normative, mais d'une manière indirecte. Pour s'en rendre compte, nous devons regarder le débat actuel à propos de la méthodologie en éthique normative.

Un thème dominant en éthique normative pendant le siècle dernier au moins a été le débat entre ceux qui soutiennent une théorie éthique normative systématique – l'utilitarisme et d'autres formes de conséquentialisme ont été de ce côté les principaux concurrents – et ceux

²³² Darwin à Charles Lyell, in F. Darwin (éd.), *The Life and Letters of Charles Darwin*, Volume II, Londres, Murray, 1887, p. 262.

²³³ Voir, par exemple, E. O. Wilson, *On Human Nature*, Cambridge, Harvard University Press, 1978, p. 5.

qui fondent leur éthique normative sur nos jugements moraux ou nos intuitions morales. Dans ce débat, les armes principales des adversaires de l'utilitarisme ont été des exemples visant à montrer que les injonctions de l'utilitarisme se heurtent aux intuitions morales que nous partageons tous. L'exemple littéraire le plus connu se trouve peut-être dans *Les Frères Karamazov*, où Dostoïevski montre Ivan mettant Aliocha au défi de dire s'il consentirait à bâtir un monde dans lequel les gens seraient heureux et en paix, si ce monde idéal ne pouvait être accompli qu'en torturant « ce petit enfant frappant sa poitrine de ses petits poings ». Aliocha répond qu'il ne consentirait pas à fonder un tel monde sur ces conditions²³⁴. Hasting Rashdall pensait qu'il pouvait réfuter l'utilitarisme hédoniste en soutenant qu'il ne pouvait expliquer la valeur de la pureté sexuelle²³⁵. H. J. McCloskey, écrivant à une époque où les lynchages dans le sud des États-Unis étaient encore possibles, pensait que le fait que la théorie pouvait imposer à un shérif d'incriminer un homme innocent pour empêcher une foule de blancs de lyncher une demi-douzaine d'innocents en représailles d'un viol était une objection décisive à l'utilitarisme²³⁶. Bernard Williams a proposé l'exemple similaire d'un botaniste arrivant dans un village dans la jungle où vingt innocents sont sur le point d'être fusillés. On lui dit que dix-neuf d'entre eux seront épargnés, à condition qu'il abatte lui-même le vingtième. Bien que Williams lui-même n'ait pas dit qu'il serait nécessairement immoral d'abattre le vingtième, il pensait que l'utilitarisme ne pouvait pas rendre compte de la difficulté de la décision²³⁷.

Initialement, l'usage de tels exemples pour faire appel à nos intuitions morales ordinaires contre les théories conséquentialistes était un expédient *ad hoc* sans fondements métaéthiques. C'était simplement une manière de dire : « si la théorie U est vraie, alors dans la situation X il faudrait faire Y. Mais nous savons qu'il serait mal de faire Y dans la situation X, par conséquent U ne peut pas être vraie. » C'est un argument efficace contre U, aussi longtemps que le jugement selon lequel il serait mal de faire Y en situation X n'est pas remis en question. Mais l'argument ne fait rien pour établir qu'il serait mal de faire Y dans la situation X, ni pour

²³⁴ F. Dostoïevski, *Les Frères Karamazov*, Henri Mongault (tr. fr.), Paris, Gallimard, 1994, Partie 2, Livre 5, Chapitre 4, p. 339.

²³⁵ H. Rashdall, *The Theory of Good and Evil*, Volume 1, Oxford, Clarendon Press, 1907, p. 197.

²³⁶ H. J. McCloskey, « An Examination of Restricted Utilitarianism », in M. D. Bayles (éd.), *Contemporary Utilitarianism*, Gloucester, Peter Smith, 1978, l'exemple se trouve p. 121.

²³⁷ B. Williams, « A Critique of Utilitarianism », in J. J. C. Smart et B. Williams, *Utilitarianism; For and Against*, Cambridge, Cambridge University Press, 1973, p. 96-100, 110-117.

indiquer de quelle nature serait une théorie plus solide que U. Dans sa *Théorie de la justice*, John Rawls a fait un pas décisif vers la fusion de cet argument avec une méthodologie éthique lorsqu'il a soutenu que l'épreuve de vérité d'une théorie morale solide est qu'elle puisse atteindre un « équilibre réfléchi »²³⁸ (*reflective equilibrium*) avec nos jugements moraux. Par « équilibre réfléchi », Rawls voulait dire que là où il n'y pas de théorie intrinsèquement convaincante qui corresponde parfaitement à nos jugements moraux initiaux, nous devrions modifier soit la théorie, soit les jugements, jusqu'à ce que nous obtenions un équilibre entre les deux. Le modèle en l'occurrence est celui de l'évaluation d'une théorie scientifique. En science, nous acceptons généralement la théorie qui correspond le mieux aux données, mais quelquefois, si la théorie est convaincante en soi, nous pouvons être prêts à l'accepter même si elle ne les intègre pas toutes. Nous pouvons supposer que les données écartées sont erronées, ou qu'il existe des facteurs encore inconnus à l'œuvre dans cette situation particulière. Dans le cas d'une théorie éthique normative, suppose Rawls, les données brutes sont nos jugements moraux préalables. Nous essayons de les assortir à une théorie convaincante, mais si nous n'y arrivons pas, nous rejetons certains de ces jugements et modifions la théorie pour qu'elle corresponde aux autres. Finalement, la correspondance entre une théorie et les jugements restants atteint un équilibre et nous avons alors la meilleure théorie possible. Sous cet angle l'acceptation d'une théorie morale n'est pas déterminée par la cohérence interne et la plausibilité de la théorie en elle-même, mais, dans une large mesure, par son accord avec les jugements moraux préalables que nous refusons de réviser ou d'abandonner. Dans la *Théorie de la justice*, Rawls utilise ce modèle pour justifier le rafistolage de son idée originale d'un choix résultant d'un contrat hypothétique, jusqu'à ce qu'il soit capable de produire des résultats moins éloignés de nos idées ordinaires de justice²³⁹.

Le modèle de l'équilibre réfléchi m'a toujours semblé douteux. L'analogie entre le rôle d'une théorie morale normative et d'une théorie scientifique est fondamentalement inadaptée²⁴⁰. Une théorie scientifique cherche à expliquer l'existence de données qui renvoient à un monde « là-

²³⁸ Le terme a également été régulièrement traduit par *équilibre réflexif*. Nous reprendrons cependant ici une traduction de Catherine Audard (voir note 27), dans la mesure où parler d'équilibre *réfléchi* souligne la dimension de délibération (individuelle ou collective) inhérente à la méthode de Rawls et à laquelle s'intéresse Singer (N.d.T.).

²³⁹ J. Rawls, *Théorie de la justice*, Paris, Points, 2009, p. 47-48. L'idée d'un équilibre réfléchi était déjà présente in J. Rawls « Outline of a Decision Procedure for Ethics », *The Philosophical Review*, 60, 1951, p. 177-197. L'analogie avec une théorie scientifique est explicite dans l'article précédent.

²⁴⁰ Voir P. Singer, « Sidgwick and Reflective Equilibrium ».

dehors » que nous cherchons à expliquer. Admettons que les données puissent avoir été affectées par des erreurs de mesure ou d'interprétation ; à moins de pouvoir rendre compte de ce qu'ont pu être ces erreurs, ce n'est pas à nous de choisir ou de rejeter les observations. Une théorie éthique normative, cependant, n'essaie pas d'expliquer nos intuitions morales ordinaires. Elle pourrait en rejeter l'intégralité, et pourtant rester supérieure aux autres théories normatives qui correspondaient mieux à nos jugements moraux. Car une morale normative n'est pas une tentative de répondre à la question « Pourquoi pensons-nous comme nous le faisons en matière de morale ? ». Même sans une compréhension évolutive de l'éthique, il est évident que la question « Pourquoi pensons-nous comme nous le faisons en matière de morale ? » peut exiger une enquête historique, plutôt que philosophique. À propos de l'avortement, du suicide et de l'euthanasie volontaire par exemple, nous pouvons penser comme nous le faisons parce que nous avons grandi dans une société qui a été, pendant près de 2000 ans, dominée par la religion chrétienne. Nous pouvons désormais ne plus croire au christianisme comme autorité morale, mais nous pouvons trouver difficile de nous débarrasser de nos intuitions morales, formées par nos parents et nos professeurs, qui étaient eux-mêmes croyants, ou étaient influencés par d'autres qui l'étaient.

Une théorie morale normative est une tentative de répondre à la question « Que devrions nous faire ? ». Il est parfaitement possible de répondre à cette question en disant : « Il nous faut ignorer tous nos jugements moraux ordinaires, et faire ce qui produira les meilleures conséquences. » Bien sûr, on pourrait avoir besoin de donner une sorte d'argument pour cette réponse. Mon propos à présent n'est pas de fournir cet argument, ni aucun autre argument en faveur d'alternatives possibles à une théorie quelconque qui expliquerait le mieux nos jugements intuitifs. Ce qui m'importe est que le modèle de l'équilibre réfléchi – du moins tel qu'il est présenté dans la *Théorie de la justice* – semble exclure une telle réponse, parce qu'il suppose que nos intuitions morales sont des sortes de données à partir desquelles nous pouvons apprendre ce que nous devrions faire.

Rawls a de nouveau décrit les implications métaéthiques de sa méthode dans son ouvrage *Libéralisme politique*²⁴¹. Il y a distingué sa méthode d'un intuitionnisme éthique démodé, en le décrivant à la place comme « constructivisme kantien ». Alors que l'intuitionnisme cherche à défendre nos intuitions en tant qu'elles offrent une compréhension rationnelle de vrais principes éthiques, le constructivisme remplace cette défense par une recherche de « motifs raisonnables permettant d'atteindre

²⁴¹ J. Rawls, *Political Liberalism*, New York, Columbia University Press, 1993.

un accord, enracinés dans notre conception de nous-mêmes et dans notre relation à la société ». Nous ne pouvons pas, sous cet angle, découvrir de vérité morale. Nous pouvons seulement construire nos conceptions morales à partir de concepts et d'idées que nous possédons déjà.

Une objection évidente au constructivisme kantien de Rawls est qu'il rend l'éthique culturellement relative. Des personnes différentes, avec des conceptions d'elles-mêmes et de leur relation à la société différentes, pourraient construire des théories différentes qui les mèneraient à des principes de justice différents. Le cas échéant, on ne pourrait alors pas dire qu'un ensemble de principes est vrai et l'autre faux. Le maximum que l'on puisse accorder aux principes particuliers de justice que défend Rawls, c'est qu'ils fournissent des motifs raisonnables permettant un accord à ceux qui partagent « notre » conception de nous-mêmes et de notre relation à la société. Mais certains ne verront pas ceci comme une objection. Le relativisme culturel a toujours eu de nombreux partisans en éthique, dont beaucoup croient à tort qu'il offre une défense contre l'impérialisme culturel. (C'est le contraire de la vérité. Si l'éthique est culturellement relative, et que ma culture donne une grande importance à l'imposition de nos propres valeurs aux autres cultures, le relativisme éthique ne donne aucun appui pour affirmer que nous sommes dans l'erreur en croyant qu'il est bon d'imposer nos valeurs aux autres.) Je ne veux pas, cependant, m'éterniser sur les éléments relativistes du constructivisme kantien, car je compte faire une objection plus générale à toute manière de faire de l'éthique qui juge une théorie normative soit dans son entier, soit en partie, par la mesure à laquelle celle-ci correspond à nos intuitions morales.

J'admets volontiers qu'il est possible d'interpréter le modèle de l'équilibre réfléchi de telle manière qu'il prenne en compte n'importe quel motif afin de s'opposer à nos intuitions, y compris ceux que j'ai pu avancer. Norman Daniels a défendu de façon persuasive cette interprétation « large » de l'équilibre réfléchi²⁴². Si l'interprétation est effectivement suffisamment large pour approuver le rejet de toutes nos croyances morales ordinaires, alors je n'y oppose pas d'objection. Cependant, le prix à payer pour éviter le conservatisme inhérent à l'interprétation étroite est que l'équilibre réfléchi cesse d'être une méthode distinctive en éthique normative. Alors qu'il y avait auparavant un contraste entre la méthode de l'équilibre réfléchi et les tentatives « fondationnalistes » de construire un système éthique à partir de quelque point de départ indubitable, le fondationnalisme devient maintenant simplement le cas limite d'un équilibre réfléchi large.

²⁴² Voir N. Daniels, *Justice and Justification: Reflective Equilibrium in Theory and Practice*, Cambridge, Cambridge University Press, 1997.

Revenons un instant aux situations du dilemme du tramway. Comme mentionné auparavant, les examens philosophiques de ces situations à partir de la contribution de Thomson ont été motivés par la recherche de différences entre les situations qui puissent justifier nos réactions intuitives initiales. Cependant, si Greene a raison de suggérer que nos réactions intuitives sont dues à des différences de charge émotionnelle entre des situations qui impliquent de provoquer la mort de quelqu'un d'une manière rapprochée, personnelle, et des situations qui entraînent la mort de cette même personne d'une manière distanciée et moins personnelle, pourquoi devrions nous croire qu'il existe quelque chose qui justifie ces réactions ? Si les résultats initiaux de Greene sont confirmés par des recherches ultérieures, nous pourrions au bout du compte conclure qu'il a non seulement expliqué, mais qu'il a supprimé le casse-tête philosophique (Je dis que nous pourrions parvenir à cette conclusion au bout du compte, car les données seules de Greene ne peuvent bien entendu pas prouver la vérité ou la fausseté de conceptions normatives. Un argument normatif est requis, dont j'esquisserai plus bas la nature, pour relier ces données à une conception normative particulière.).

La question devient plus claire dès lors que nous prenons en considération à quel point les conclusions de Greene s'intègrent à la conception évolutive plus large des origines de la moralité énoncées précédemment²⁴³. Pour la plus grande part de notre histoire évolutive, les êtres humains ont vécu en de petits groupes, et il est presque certain qu'il en est allé de même pour nos ancêtres, primates pré-humains et mammifères sociaux. Dans ces groupes, la violence ne pouvait être infligée que de manière rapprochée et personnelle – en frappant, poussant, étranglant, ou en utilisant un bâton ou une pierre comme massue. Pour composer avec de telles situations, nous avons développé des réactions immédiates et reposant sur des émotions, permettant de répondre à des questions impliquant des interactions rapprochées et personnelles avec les autres. La pensée de pousser un inconnu du haut d'une passerelle suscite ces réactions émotives. Tirer un levier qui déviera un train qui lui-même heurtera quelqu'un n'entretient de ressemblance avec aucune action susceptible d'avoir eu lieu dans les circonstances dans lesquelles nous et nos ancêtres avons vécu. Par conséquent, la pensée de tirer le levier ne suscite pas la même réaction émotionnelle que la pensée de pousser quelqu'un du haut d'une passerelle. Ainsi, le trait marquant expliquant la différence de jugements intuitifs dans les deux cas est que le cas de la passerelle renvoie à un type de situation susceptible de s'être présenté au cours de l'éternité durant laquelle nous

²⁴³ Comme Greene lui-même l'a souligné. Voir note 14, ci-dessus.

étions en train d'évoluer ; alors que le problème standard du tramway décrit une manière de provoquer la mort de quelqu'un qui n'a été possible qu'au cours de ce siècle ou des deux derniers, un intervalle de temps bien trop court pour avoir un impact sur nos modèles héréditaires de réaction émotionnelle. Mais quelle est l'importance morale du fait que j'ai tué quelqu'un d'une manière qui était déjà possible il y a un million d'années, plutôt que d'une manière rendue possible il y a seulement deux-cents ans ? Je répondrais : aucune.

Les récents progrès scientifiques évoqués ont ainsi une certaine portée normative, et à différents niveaux. Au niveau particulier de l'analyse de problèmes moraux comme ceux posés par les incidents de tramway, une meilleure compréhension de la nature de nos réactions intuitives suggère qu'il ne sert à rien d'essayer de trouver des principes moraux qui justifient les intuitions différentes que provoquent les diverses expériences de pensée. Il n'existe très probablement pas de distinction moralement pertinente entre les situations. Au niveau plus général de la méthode en éthique, cette même compréhension de la manière selon laquelle nous formulons des jugements moraux jette un sérieux doute sur la méthode de l'équilibre réfléchi. Il y a peu d'intérêt à construire une théorie morale conçue pour correspondre à des jugements moraux, si ces jugements proviennent eux-mêmes de nos réactions évoluées à des situations dans lesquelles nous et nos ancêtre avons vécu au cours des périodes de notre évolution comme mammifères sociaux, comme primates, et enfin comme êtres humains. Nous devrions, avec nos facultés actuelles de raisonnement et l'évolution rapide de la situation dans laquelle nous vivons, être capables de faire mieux que cela.

Un défenseur de l'idée de l'équilibre réfléchi pourrait dire que ces arguments contre le fait de donner du poids à certaines intuitions peuvent eux-mêmes, sur le modèle de « l'équilibre réfléchi élargi », faire partie du processus réalisant un équilibre entre une théorie et nos jugements moraux. Les arguments nous conduiraient alors à rejeter les jugements que nous pourrions autrement conserver, et nous feraient ainsi aboutir à une théorie normative différente. Comme nous l'avons déjà constaté, rendre le modèle de « l'équilibre réfléchi » aussi universel peut le rendre récupérable, mais le prix à payer est alors de le rendre presque inepte. Car avec ce changement, les « données » auxquelles une théorie morale solide est supposée correspondre sont devenues si changeantes qu'elles ne peuvent jouer, au mieux, qu'un rôle mineur dans la détermination de la forme finale de la théorie morale normative. Finalement, pour les mêmes raisons qui font que l'équilibre réfléchi ne peut plus se présenter comme une manière d'évaluer une théorie morale, le constructivisme kantien cesse d'être une métaéthique

attrayante, qu'il finisse ou non par être culturellement relatif. Dans la mesure où « notre conception de nous-mêmes » est liée à nos idées intuitives de bon et de mauvais, nous pouvons nous demander pourquoi nous devrions nous soucier de construire une conception morale portant sur la bonne manière d'agir dans des situations particulières à partir de nos intuitions évoluées. De plus, un constructiviste kantien qui réussirait à éviter le relativisme culturel en trouvant des idées intuitives du bien et du mal universellement partagées peut n'avoir rien dégagé d'autre que ce que notre héritage évolutif commun nous a, sans surprise, légué comme ensemble commun d'idées intuitives du bien et du mal.

En bref, je dis ceci. Les progrès en ce qui concerne notre compréhension de l'éthique n'impliquent pas eux-mêmes directement de conclusions normatives, mais ils sapent certaines conceptions quant à la manière de faire de l'éthique, qui ont elles-mêmes des conclusions normatives. Ces conceptions de l'éthique ont tendance à être trop respectueuses de nos intuitions. Notre meilleure compréhension de l'éthique nous donne des raisons d'être moins respectueux à leur égard.

V. Conclusion : une voie à suivre ?

Chaque fois qu'il est suggéré que l'éthique normative ne doit pas tenir compte de nos intuitions morales ordinaires se pose l'objection selon laquelle sans ces intuitions, nous ne pouvons aller nulle part. Il y a eu au cours des siècles de nombreuses tentatives pour trouver des principes premiers en éthique, mais la plupart des philosophes considère qu'elles ont toutes échoué. Même une théorie éthique radicale telle que l'utilitarisme doit se reposer sur une intuition fondamentale de ce qui est bon. Ainsi, il semble que nous soyons laissés à nos intuitions, sans rien de plus. Si nous les rejetons toutes, nous devons devenir des sceptiques ou des nihilistes éthiques.

On pourrait essayer de répondre à cette objection de bien des manières, et je n'ai pas le temps de les passer ici en revue. Laissez-moi ainsi juste suggérer une possibilité. La recherche comportementale de Haidt et les études d'imagerie cérébrale de Greene suggèrent la possibilité de distinguer nos réactions émotives immédiates de nos conclusions plus raisonnées. Dans la vie quotidienne, comme le signale Haidt, notre faculté de raisonner est susceptible de n'être rien de plus qu'une rationalisation de nos réactions intuitives – comme il le formule, le chien émotionnel fait remuer sa queue

rationnelle. Mais les recherches de Greene suggèrent que chez certaines personnes, le raisonnement peut surmonter une réaction intuitive initiale. Ceci, au moins, semble la manière la plus convaincante de rendre compte du temps de réaction plus élevé chez les sujets qui, dans l'exemple de la passerelle, ont conclu qu'il serait justifié de pousser l'inconnu sous le tramway. Ces personnes semblent avoir eu les mêmes réactions émotionnelles opposées à l'idée de pousser l'inconnu, mais une réflexion plus poussée les a conduites à rejeter cette réaction émotive et à donner une réponse différente. Les données provisoires montrant une activité plus importante dans les parties de leur cerveau associées aux processus cognitifs suggèrent la même conclusion. De plus, la réponse donnée par ces sujets est certainement la réponse rationnelle. La mort d'une personne est une tragédie moindre que la mort de cinq personnes. Ce raisonnement nous conduit à tirer le levier dans le cas standard du tramway, et il devrait également nous conduire à pousser l'inconnu de la passerelle, dans la mesure où il n'existe pas de différences morales pertinentes entre les deux situations (même si nous pouvons décider de refuser de faire l'éloge de gens capables de pousser quelqu'un d'une passerelle dans ces circonstances. Comme le pointe Henry Sidgwick dans *Les Méthodes de l'éthique*, il est important de faire une distinction entre l'utilité d'une action, et l'utilité de louer ou de blâmer cette action. Nous pouvons ne pas souhaiter louer les personnes capables de pousser des inconnus dans le vide, de peur qu'elles ne le fassent en d'autres occasions lorsque cela ne sauvera pas plus de vie que cela n'en coûte²⁴⁴).

On pourrait dire que la réponse que j'ai appelée « plus raisonnée » reste cependant fondée sur une intuition, par exemple l'intuition que causer cinq morts est pire qu'en causer une seule, ou plus fondamentalement, l'intuition que le fait qu'une personne soit tuée est une mauvaise chose. Mais si c'est une intuition, elle est différente de l'intuition à laquelle Haidt et Greene se réfèrent. Elle ne semble pas faire partie des intuitions résultant de notre passé évolutif. Nous avons déjà énoncé l'observation de Hume selon laquelle « il n'y a pas dans l'esprit de l'homme une passion telle que l'amour de l'humanité, uniquement comme telle » et comme nous l'avons vu, il y a une bonne raison évolutive de la raison pour laquelle il devrait en être ainsi. Par conséquent, l'« intuition » qui nous dit que la mort d'une personne est une tragédie moindre que la mort de cinq personnes n'est pas semblable aux intuitions qui nous disent qu'on peut tirer le levier, mais pas pousser l'inconnu de la passerelle. Il peut être plus proche de la vérité de

²⁴⁴ H. Sidgwick, *The Methods of Ethics*, Seventh Edition, London, Macmillan, 1907, p. 428-429.

dire que c'est une intuition rationnelle, quelque chose de semblable aux trois « axiomes éthiques » ou aux « intuitions éthiques vraiment claires et certaines »²⁴⁵ auxquels Henry Sidgwick fait référence dans sa défense de l'utilitarisme dans *Les Méthodes de l'éthique*. Le troisième de ces axiomes est que « le bien d'un individu n'a pas plus d'importance, du point de vue de l'univers (si je puis dire), que le bien d'un autre individu »²⁴⁶.

Peut-être qu'ici, après nous être trouvés largement d'accord avec Hume pour la plus grande partie de cet article, nous éprouvons le besoin d'en appeler à quelque chose qui se trouve chez un penseur aux antipodes de Hume, à savoir Emmanuel Kant. Kant pensait qu'à moins que la moralité puisse être fondée sur la raison pure, elle n'était qu'une chimère²⁴⁷. Peut-être avait-il raison. À la lumière des meilleures compréhensions scientifiques de l'éthique, nous faisons face à un choix. Nous pouvons adopter la conception selon laquelle nos intuitions et nos jugements moraux sont et seront toujours des réactions émotives, et selon laquelle la raison ne peut rien faire d'autre que construire le meilleur argument possible pour une décision déjà prise sur un terrain non rationnel. Cette approche conduit à une forme de scepticisme moral, un scepticisme certes toujours compatible avec la défense de nos valeurs morales émotionnellement fondées et avec une incitation à avoir les idées claires à leur sujet²⁴⁸. Sinon, nous pourrions nous consacrer à la tâche ambitieuse de séparer les jugements moraux que nous devons à notre histoire évolutive et culturelle de ceux qui ont une base rationnelle. C'est là une tâche vaste et difficile. Préciser dans quel sens un jugement moral peut avoir une base rationnelle n'est pas simple. Cela me semble néanmoins en valoir la peine, dans la mesure où c'est la seule façon d'éviter le scepticisme moral.

²⁴⁵ *Idem*, partie III, chapitre XIII, §5 (N.d.T.).

²⁴⁶ *Idem*, p. 382.

²⁴⁷ E. Kant, *Fondation de la métaphysique des mœurs*, Alain Renaut (tr. fr.), Paris, Flammarion, 1994, Section II.

²⁴⁸ Greene adopte cette position dans *The Terrible, Horrible, No Good, Very Bad Truth about Morality*, et dans *What to Do About It*. Il définit sa conception comme un scepticisme moral, mais la distingue du nihilisme moral, dans lequel il n'y a pas de place du tout pour des valeurs morales. Je lui suis reconnaissant, non seulement pour ses recherches éclairantes, mais également pour ses commentaires précieux sur cet article. D'autres commentaires utiles sont venus de la part de personnes trop nombreuses pour être nommées individuellement, je remercie donc collectivement tous ceux qui intervinrent lorsque j'ai présenté ce texte à la James Rachels Memorial Conference à l'Université de l'Alabama ; au Princeton University Center for Human Values Fellows' Seminar ; et aux départements de des universités suivantes : Université de Melbourne, Université du Vermont, Université Rutgers, et Université de Lodz. En dépit de tous ces conseils, je suis conscient qu'un travail bien plus grand est requis ; cet article n'est rien de plus que l'esquisse d'un argument que j'espère développer de manière plus satisfaisante à l'avenir.