



**HAL**  
open science

# De l'histoire de la pensée économique à la philosophie morale

Christophe Salvat

► **To cite this version:**

Christophe Salvat. De l'histoire de la pensée économique à la philosophie morale : Perspectives historiques et analytiques. Economies et finances. Aix-Marseille Université, 2020. tel-03083701

**HAL Id: tel-03083701**

**<https://shs.hal.science/tel-03083701>**

Submitted on 19 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Dossier en vue de l'obtention de l'Habilitation à Diriger les Recherches

De

**M. Christophe Salvat**

Le 15 octobre 2020

Aix-Marseille Université

**Membres du jury :**

- M. Pascal Taranto, AMU (garant)
- Mme Carla Bagnoli, Université de Modène, Italie (rapporteur)
- Mme Herrade Igersheim, CNRS, université de Strasbourg (rapporteur)
- M. Emmanuel Picavet, Université Paris I Panthéon-Sorbonne (rapporteur)
- M. Laurent Jaffro, Université Paris I Panthéon-Sorbonne



# Table des matières

## **1. Présentation analytique de mes recherches**

*De l'histoire de la pensée économique à la philosophie morale : Perspectives historiques et analytiques*

## **2. Sélection de travaux (I)**

*L'utilitarisme*, ouvrage en cours de publication, La découverte, collection Repères, 2020.

## **3. Sélection de travaux (II)**

Sélection d'articles publiés entre 2006 et 2020



Aix-Marseille Université (AMU)

**Présentation analytique de mes recherches dans la perspective de**

l'obtention d'une Habilitation à Diriger les Recherches

*De l'histoire de la pensée économique à la philosophie*

*morale :*

*Perspectives historiques et analytiques*

Christophe Salvat

CNRS, Centre Gilles Gaston Granger UMR 7304



# Table des matières

<b>INTRODUCTION</b> .....	9
<b>Première Partie</b> .....	15
<b>De la scientificité du discours économique : approches historiques</b> .....	15
<b>1. ANDRE MORELLET : FORMATION ET DIFFUSION DE LA PENSEE ECONOMIQUE LIBERALE FRANÇAISE</b> .....	16
A. METHODE ET OBJET .....	18
B. LOBBYING ET EXPERTISE ECONOMIQUE AU XVIIIIE SIECLE .....	19
C. DEFINIR (ET IMPOSER) UNE DEFINITION DE L'ECONOMIE POLITIQUE .....	25
<b>2. L'ECONOMIE POLITIQUE DES LUMIERES OU LE COMBAT DES ENCYCLOPEDISTES</b> .....	28
A. L'ECONOMIE POLITIQUE, NOUVELLE BRANCHE DES SAVOIRS DANS L'ENCYCLOPEDIE DE DIDEROT ET D'ALEMBERT .....	30
B. LA CONCEPTION ALTERNATIVE DE L'ECONOMIE POLITIQUE CHEZ JEAN-JACQUES ROUSSEAU .....	36
<b>3. L'ECONOMIE ET LA LOI CHEZ ROUSSEAU</b> .....	40
A. L'ECONOMIE ET LA LOI CHEZ ROUSSEAU .....	42
B. LA RECEPTION DE ROUSSEAU .....	50
<b>Deuxième partie</b> .....	59
<b>Doutes méthodologiques et réorientation disciplinaire</b> .....	59
<b>1. De l'histoire intellectuelle et de ses limites</b> .....	60
A) HISTOIRE DE LA PENSEE ECONOMIQUE ET HISTOIRE INTELLECTUELLE .....	60
B) LIMITES DE L'HISTOIRE INTELLECTUELLE .....	63
C) LES RAISONS D'UNE REORIENTATION .....	70
<b>2) De la philosophie économique</b> .....	74
A) CE QU'EST LA PHILOSOPHIE ECONOMIQUE .....	75
B) LE PATERNALISME COMPORTEMENTAL .....	82
C) THEORIE REDUCTIONNISTE DE L'IDENTITE PERSONNELLE .....	92
<b>TROISIEME PARTIE</b> .....	109
<b>LES TROIS NIVEAUX DE L'ETHIQUE</b> .....	109
<b>1. Ethique normative et utilitarisme</b> .....	110
A) LE PRINCIPE D'IMPARTIALITE .....	111
B) L'UTILITARISME ET LA REGLE .....	119



<b>2. Méta-éthique et réalisme moral</b> .....	134
A) CADRE META-ETHIQUE : LE FONDAMENTALISME DES RAISONS .....	135
B) DE L'IMPACT DE LA META-ETHIQUE SUR L'ETHIQUE NORMATIVE : L'EXEMPLE DE LA TRIPLE THEORIE DE PARFIT .....	144
<b>3. Ethique appliquée</b> .....	151
A) ETHIQUE APPLIQUEE ET THEORIE NORMATIVE .....	152
B) L'ETHIQUE ANIMALE .....	156
C) LE RECHAUFFEMENT CLIMATIQUE.....	160
<b>QUATRIEME PARTIE</b> .....	163
<b>ANIMATION DE LA RECHERCHE ET CV</b> .....	163
<b>1. Animation de la recherche</b> .....	164
<b>2. Curriculum Vitae</b> .....	169
<b>Bibliographie</b> .....	185

# INTRODUCTION

Ce mémoire présente et discute environ vingt ans de recherches menées respectivement à Lyon (au Centre Auguste et Léon Walras), à Dublin (à Trinity College), à Lausanne (au Centre Walras-Pareto), à Aix-en-Provence (au GREQAM, UMR 6579), à Cambridge (au Centre for History and Economics), de nouveau à Lyon (à Triangle, UMR 5206), puis à Aix-en-Provence (Centre Gilles-Gaston Granger UMR 7304).

De retour dans la ville où j'ai débuté ma carrière au CNRS, mais cette fois dans un laboratoire de philosophie, il m'a semblé qu'il était temps pour moi de me solliciter une Habilitation à Diriger les Recherches, et de clore ainsi un parcours intellectuel et scientifique qui a débuté en histoire de la pensée économique (et politique), qui m'a conduit à la philosophie économique, et qui trouve aujourd'hui son aboutissement en philosophie.

Cette demande d'habilitation répond à deux objectifs, le premier étant, naturellement, de pouvoir encadrer des thèses de doctorat. Au-delà de l'intérêt et de la satisfaction personnelle que cette fonction peut apporter, la possibilité de diriger de thèse me permettrait également de développer la philosophie économique au sein de l'axe 'Histoire de la philosophie et ontologies du présent' du centre Gilles-Gaston Granger que j'ai intégré en mars 2018.

Mais l'HDR ne se réduit pas à cette possibilité offerte, aussi importante soit-elle sur le plan institutionnel. Elle est également, ou plus exactement peut être vue comme, un moment de réflexion et de mise en perspective de ses travaux, passés et à venir. Dans mon cas, cela constitue un aspect déterminant de ma démarche, et c'est la raison pour laquelle je n'ai

pas souhaité engager cette réflexion plus tôt. La distance qui me sépare aujourd'hui de mes travaux de thèse, mais également l'évolution institutionnelle, géographique et disciplinaire qui a été la mienne, me permettent aujourd'hui de reconsidérer l'ensemble de mon parcours et d'en isoler, malgré les changements de thématique, une ligne directrice constante.

Mon parcours personnel peut surprendre au premier abord, d'historien de la pensée économique au dix-huitième siècle, je me suis spécialisé sur les questions de rationalité, de choix inter-temporels et d'éthique. Cette évolution, bien que radicale en certains aspects, n'a pas été totalement arbitraire. Elle est le produit, en fait, d'un questionnement sur la normativité de la science économique, demeuré sensiblement identique, mais dont les formes ont varié et évolué au gré des mutations de la discipline elle-même. La présentation générale de mes travaux me donne aujourd'hui l'opportunité de mettre en perspective cette évolution et de lui donner la cohérence dont elle semble manquer.

J'ai souhaité la présenter le plus simplement possible, de façon (grossièrement) chronologique. Mon parcours se divise en trois grands 'moments'. Le premier moment correspond à mes travaux d'histoire de la pensée économique réalisés pendant et immédiatement après ma thèse. Ils s'inscrivent dans la tradition de l'histoire intellectuelle incarnée notamment par Jean-Claude Perrot. Ils correspondent assez bien au type de travaux qui était alors en cours au Centre Auguste et Léon Walras, et pour lesquels ce centre de recherche avait acquis une certaine notoriété. Une première partie de ces travaux est issue de ma thèse et porte sur la contribution d'André Morellet au développement et à la diffusion du discours économique en France durant la seconde moitié du dix-huitième siècle. La seconde partie, que j'ai développée à mon arrivée au

Greqam en tant que chargé de recherche CNRS, porte sur la représentation de l'économie politique dans l'*Encyclopédie* de Diderot et d'Alembert. Je l'ai poursuivi lors d'un séjour long à l'université de Cambridge au sein du département d'histoire<sup>1</sup>, où enseignaient alors les principaux tenants de l'histoire intellectuelle anglaise, à savoir Quentin Skinner, John Dunn et Isvant Hont<sup>2</sup>.

C'est, paradoxalement, pendant ce séjour de recherche et d'enseignement (j'étais *associate lecturer* à la faculté d'histoire et de philosophie) que mon rapport à l'histoire et aux lectures contextualisées a totalement changé, et que j'ai décidé de me rapprocher de la philosophie. J'ai alors investi le relativement nouveau champ de la philosophie économique, développé entre autres par Alain Leroux au Greqam. Je me suis alors intéressé à la question du paternalisme libertarien de Sunstein et Thaler, et plus généralement à ce que l'on appelle la philosophie économique. L'émergence et le succès de la théorie dite des 'nudges' ou du paternalisme libertarien au début des années 2000 a été l'occasion pour moi d'interroger la notion de normativité du standard de rationalité économique, autrement dit de questionner le fondement normatif des normes économiques et leur articulation avec les normes morales. Les travaux de Sunstein et Thaler, dont la dimension théorique est cependant relativement décevante, m'ont amené à reconsidérer mon

---

<sup>1</sup> Je suis resté à Cambridge de 2006 à 2011, en tant que *Research Fellow* à Robinson College. Je fus admis au *Centre for History and Economics* dirigé par Emma Rothschild et Gareth Stedman Jones, en tant que *visiting scholar*. C'est grâce notamment au soutien financier et logistique de ce centre, que j'ai pu organiser en 2008 (14 et 15 juillet) un colloque inter-disciplinaire sur le thème de la citoyenneté à Cambridge. Une sélection de ces travaux a été publiée dans *History of European Ideas*. C. Salvat (ed.), "Rethinking Citizenship", *History of European Ideas*, Volume 37, Issue 2, June 2011.

<sup>2</sup> J.G.A. Pocock, *The Machiavellian Moment: Florentine Political Thought and the Atlantic Republican Tradition*, Princeton: Princeton University Press, 1975; Q. Skinner, *The Foundations of Modern Political Thought*, Cambridge: Cambridge University Press, 1978; J. Dunn, *The Political Thought of John Locke: An Historical Account of the Argument of the 'Two Treatises of Government'*, Cambridge: Cambridge University Press, 1982.

parcours de chercheur et à me réorienter vers la discipline philosophique. Je suis, en effet, arrivé à la conclusion que, si les recherches interdisciplinaires doivent être encouragées, celles-ci ne doivent cependant pas se traduire par une négation de l'expertise disciplinaire. En d'autres termes, faire de la philosophie économique ne peut se réduire à investir des problématiques philosophiques avec un cadre théorique économique (ou inversement). La question de la norme, et notamment de la norme morale, étant spécifiquement une question philosophique, j'ai demandé, tout d'abord, à rejoindre le laboratoire multidisciplinaire Triangle à l'Ecole Normale Supérieure de Lyon, puis le Centre Gilles Gaston Granger d'Aix-en-Provence.

Depuis, mes travaux portent essentiellement sur des sujets à forte dimension philosophique tels les notions de rationalité, d'identité personnelle ou d'éthique. Je me suis ainsi spécialisé sur la pensée du philosophe britannique Derek Parfit, notamment sur sa conception réductionniste de l'identité personnelle, sur son rapport à l'utilitarisme et sur les hypothèses méta-éthiques de sa philosophie. L'approche historique, bien que présente, y occupe une place tenue. Je considère que l'étude de l'histoire de la pensée, qu'elle soit économique, philosophique ou autre, est absolument fondamentale, y compris si l'on adopte une approche a-historique. En revanche, et c'est là l'un des enseignements de l'éthique intellectuelle de Pascal Engel, cette perspective historique contextuelle ne doit pas servir de justification au relativisme scientifique et/ou moral. Or, mon expérience en histoire de la pensée a montré que c'est souvent le cas.

En parallèle à ces travaux, je développe aussi un certain nombre de questionnements sur le fondement des normes morales en éthique appliquée au regard par exemple des débats sur la Gestation pour Autrui, la génomique, l'Intelligence Artificielle ou le réchauffement

climatique. Au-delà de l'intérêt immédiat que soulève ce type de problématiques, je porte une attention toute particulière à la façon dont l'éthique appliquée nous permet de réinterroger les principes de l'éthique normative. J'en donne un exemple dans le cas de l'Intelligence Artificielle et l'agentivité morale.

On le voit, en dépit de leur grande diversité, la plupart de mes travaux ont en commun le fait d'interroger, de façons différentes, la notion de normativité. Mon travail de thèse en histoire de la pensée économique m'a, tout d'abord, conduit à envisager la question de la norme du discours économique d'un point de vue historique grâce à des travaux d'archives sur l'œuvre de l'un des premiers économistes 'professionnels' de l'histoire, l'abbé Morellet. Ces travaux avaient pour but de montrer que loin d'être animé par un désir de vérité, André Morellet travaillait avant tout pour le compte de l'administration désireuse de faire accepter ses réformes libérales par le public lettré. Ils s'inscrivent donc dans une pure perspective foucaldienne dans laquelle les normes de vérité sont avant tout des instruments de pouvoir politique. Ces travaux ont été complétés par des recherches sur l'identification et la légitimation d'un discours proprement économique au sein de l'*Encyclopédie* de Diderot et d'Alembert, ainsi que par l'étude d'un discours alternatif, celui de Rousseau, que j'ai tenté de réhabiliter. Dans le même esprit, je me suis également intéressé aux différences dans la réception de l'œuvre de Rousseau au vingtième siècle. J'ai ainsi brièvement travaillé sur les lectures maurassiennes de Rousseau et sur les attaques dont il a fait l'objet pour montrer combien le sens d'une œuvre philosophique était dépendante, non seulement du contexte socio-politique dans lequel elle a été écrite, mais également de celui dans lequel elle est lue. En dépit de l'intérêt intellectuel de ce type

d'étude, il me semble aujourd'hui qu'elles sont trompeuses. Il est vrai que le contexte socio-politique peut jouer sur la production et la réception d'une œuvre, mais cela ne remet absolument pas en cause le fait qu'il existe des normes de rationalité ou de morale, qui sont, elles, extérieures aux enquêtes dont elles font l'objet. Le comportement qui consiste à systématiquement remettre en cause l'objectivité de certaines normes, raisons ou principes conduit à relativiser non seulement leur existence mais également l'utilité de toute recherche scientifique. Or, c'est, me semble-t-il, l'un des principaux dangers qui pèsent actuellement sur notre société, et notamment dans le domaine moral. Le phénomène de réchauffement climatique constitue l'un des enjeux les plus importants que nous avons à relever dans l'histoire moderne. Mais au-delà des difficultés techniques et politiques, le principal obstacle auquel nous sommes confrontés est le relativisme scientifique. J'ai, inconsciemment, et comme beaucoup de mes collègues, participé de cette relativisation du savoir en sciences sociales. Je travaille aujourd'hui à la réhabilitation d'un réalisme moral en philosophie.

# Première Partie

## De la scientificité du discours économique : approches historiques



# 1. ANDRE MORELLET : FORMATION ET DIFFUSION DE LA PENSEE ECONOMIQUE LIBERALE FRANÇAISE

Ma démarche en histoire de la pensée économique a été beaucoup influencée par le travail que j'ai mené entre 1996 et 2001 au Centre Auguste et Léon Walras de Lyon<sup>3</sup>, au département d'économie de Trinity College à Dublin, puis au Centre Walras-Pareto de Lausanne<sup>4</sup> (2001-2002). Ayant choisi, dès la fin de maîtrise, de consacrer mon mémoire de DEA aux manuscrits de l'abbé Morellet, conservés à la Bibliothèque Municipale de Lyon, ce centre de recherche constituait naturellement l'endroit de prédilection pour poursuivre mes études de doctorat<sup>5</sup>.

---

<sup>3</sup> Le Centre Auguste et Léon Walras, aujourd'hui disparu, était rattaché au CNRS et à la faculté d'économie de Lyon 2, et spécialisé en histoire de la pensée économique. Il avait notamment bâti sa réputation sur l'édition critique des ouvrages et des manuscrits d'Auguste et Léon Walras.

<sup>4</sup> En 2001, je fus nommé Maître -assistant à l'Université de Lausanne (Suisse) et chercheur au Centre Walras-Pareto. J'enseignais l'économie politique aux premières années de droit et de sciences politiques. Côté recherche, je profitais de la présence à Lausanne des manuscrits et de la bibliothèque de Léon Walras pour initier un travail sur l'interprétation de Quesnay par le marginaliste français. Walras fut, en effet, le premier avec Marx à réaliser l'importance théorique du *Tableau économique*, qui resta ignoré la plus grande partie du dix-neuvième siècle. Cette étude me fit prendre conscience de l'importance de la réception de l'œuvre des auteurs étudiés. Mes travaux sur Rousseau me permirent d'approfondir plus tard cet aspect de l'histoire des idées.

<sup>5</sup> J'ai bénéficié lors de la rédaction de ce mémoire de thèse de l'appui financier du Ministère de l'Education et de la Recherche, qui m'a octroyé une allocation de recherche de trois ans, ainsi que l'aide de la Région Rhône-Alpes, qui grâce au projet Eurodoc, m'a permis de mener une partie de ma thèse au Trinity College de Dublin, sous la direction d'Antoin Murphy. Sur le plan intellectuel, j'ai bénéficié des conseils de mon directeur de thèse, Jean-Pierre Potier, professeur à l'université Lyon 2, mais également d'Antoin Murphy, lors de mes nombreux séjours à Dublin, ainsi que de Pierre Crépel, chercheur CNRS, qui me guida notamment dans mes travaux sur les manuscrits de Morellet.

Ma thèse portait sur l'œuvre d'André Morellet<sup>6</sup> et son influence dans le développement et la diffusion de la pensée économique en France dans la seconde moitié du dix-huitième siècle français. Le choix d'André Morellet a pu surprendre. A la différence de Quesnay ou de Turgot, Morellet n'est pas un théoricien de l'économie politique naissante, ou du moins il n'en est pas un théoricien majeur. Ses ouvrages portent essentiellement sur des questions de politiques économiques très spécifiques de l'Ancien Régime (le commerce des grains, la Compagnie des Indes, les corporations). Il est d'ailleurs surtout connu pour la vigueur – plus que pour la qualité conceptuelle – de ses pamphlets. Ce choix se justifiait, cependant, par l'approche méthodologique que j'avais choisie, à savoir celle de l'histoire intellectuelle. Je m'intéressais alors à la constitution du discours économique dans la seconde moitié du dix-huitième siècle et sur le rôle joué par l'administration politique de l'époque, notamment par Turgot. Mes travaux avaient pour ambition de mettre à jour les relations politiques qui unissaient administrateurs, ministres et pamphlétaires, dont faisait partie Morellet. Il s'agissait, en d'autres termes, de m'intéresser à la face sombre de l'émergence du discours économique, dont l'histoire de la pensée ne retient généralement que les aspects strictement intellectuels. Tous ces éléments nous invitent aujourd'hui à repenser la question de l'expert, et plus spécifiquement celle de l'expert économique, dans une société démocratique.

---

<sup>6</sup> C. Salvat, " Morellet André (1727-1819) ", in L. Frobert, J.P. Potier, A. Tiran (sous la direction de), *Economistes en Lyonnais, en Dauphiné et en Forez*, Lyon : Institut des Sciences de l'Homme, 2000, pp.228-240.

#### a. METHODE ET OBJET

La méthodologie de l'histoire de la pensée économique peut – grossièrement – se diviser en deux branches, une branche dite rétrospective et une branche dite contextuelle. L'approche rétrospective, dont Mark Blaug<sup>7</sup> fut l'une des principales figures avant de lui-même s'en détourner, consiste à considérer l'objet 'texte économique' du point de vue moderne. Appartiennent à cette branche, notamment, toutes les études généalogiques retraçant la naissance et les développements successifs d'un concept ou d'une théorie. L'approche contextuelle, au contraire, s'intéresse au texte économique en tant que manifestation ou production intellectuelle d'une époque. L'intérêt, mais également le sens, du texte n'est, selon les tenants de cette approche, accessibles qu'à ceux qui ont la connaissance requise du contexte. L'histoire intellectuelle se revendique de cette démarche<sup>8</sup>. Elle dépasse le cadre strict de l'histoire de la pensée économique. Son objet est, tout d'abord, plus large que les quelques textes classiques qui ont traversé l'histoire. Il inclut notamment tout un ensemble de textes moins connus, voire inédits, mais également la presse, les correspondances privées, les pamphlets, les affiches, les dictionnaires voire les œuvres littéraires. S'ils ne participent pas directement à l'énonciation d'une pensée, ces supports participent de sa compréhension.

De ce point de vue, le choix d'André Morellet comme sujet d'étude était totalement justifié. Bien que peu originales, les publications de Morellet donnent un remarquable aperçu des véritables enjeux, économiques mais aussi politiques, des contributions plus théoriques de Turgot ou de Quesnay notamment : la libre fabrication des toiles peintes, la suppression du

---

<sup>7</sup> M. Blaug, *Economic Theory in Retrospect*, Cambridge : Cambridge University Press, 1962-1997.

<sup>8</sup> J.C. Perrot, *Une histoire intellectuelle de l'économie politique, 17e-18e siècles*, Paris : Éditions de l'École des hautes études en sciences sociales, 1992.

privilège exclusif de la Compagnie des Indes et la libre exportation du blé, pour ne citer que ces trois-là<sup>9</sup>. La motivation de Morellet n'était pas uniquement théorique : il a été montré qu'il avait été mandaté, voire payé, par le gouvernement pour rédiger ses ouvrages. Ce fut le cas, notamment, lors de l'abolition du privilège de la Compagnie des Indes ainsi que lors de la libéralisation du commerce des grains par Turgot en 1774<sup>10</sup>.

Les manuscrits relatifs au dictionnaire sont aujourd'hui conservés à la Bibliothèque Municipale de Lyon. Ils ont été retravaillés par Morellet à la fin de sa vie, et ont généralement perdu leur forme d'article. Ce fonds comporte 92 cartons (soit environ deux milles manuscrits), et dont j'ai dressé le premier inventaire complet, reproduit en fin de thèse. La très grande diversité des sujets abordés et le nombre de mémoires rassemblés montrent l'immensité de la tâche que s'était assignée André Morellet et l'exhaustivité du dictionnaire de commerce à venir.

#### b. LOBBYING ET EXPERTISE ECONOMIQUE AU XVIIIÈME SIECLE

L'objet de ma thèse a été de montrer que le développement de l'économie politique au dix-huitième siècle fut étroitement lié aux projets de réformes économiques de l'administration. J'ai pour cela étudié le rôle joué par un économiste mineur, André

---

<sup>9</sup> C. Salvat, *Formation et diffusion de la pensée économique libérale française : André Morellet et l'économie politique du dix-huitième siècle*, Thèse de doctorat, Université Lyon 2, 2000, chapitres 1, 2 et 3.

<sup>10</sup> C. Salvat, " Un collaborateur zélé dans l'ombre de Turgot (1774-1776): l'abbé Morellet ", *Colloque International : Turgot (1727-1781), notre contemporain? Economie, administration et gouvernement au Siècle des Lumières*, Université de Caen, 14-16 mai 2003.

Morellet, membre du groupe dit de Gournay<sup>11</sup>, et proche de Turgot. La publication de sa correspondance par Dorothy Medlin et la présence de ses manuscrits à la Bibliothèque municipale de Lyon m'ont permis de remettre en perspective les enjeux politiques de la reconnaissance publique d'une science de l'économie politique.

Pour bien comprendre le rôle joué par Morellet dans le développement et la reconnaissance de l'économie politique en tant que science au dix-huitième siècle, il est nécessaire d'introduire deux personnages encore moins connus du grand public, celui de Vincent de Gournay, intendant de commerce, et celui de Daniel-Charles Trudaine, intendant des finances. Gournay se chargea de réformer les institutions économiques par l'intérieur, et délégua à d'autres la responsabilité de convaincre l'opinion du bienfait de ces réformes. Il sut s'entourer d'un certain nombre d'amis dont il dirigea les études<sup>12</sup>, et qui publièrent sous ses conseils des traductions des mercantilistes anglais. Ces publications l'aidèrent considérablement à faire accepter le mouvement de réforme dont il fut à l'origine. On peut ranger les collaborateurs de Gournay en deux groupes : 1) les administrateurs : Trudaine, Trudaine de Montigny, Maynon d'Invau, Silhouette, Malesherbes, Bouvard de Fourqueux ; 2) les écrivains : Turgot, Morellet, Dangeul, Forbonnais, Carlier, Clicquot-Blervache, Coyer..

---

<sup>11</sup> L. Charles, F. Lefebvre, C. Théré, *Le cercle de Vincent de Gournay: savoirs économiques et pratiques administratives en France au milieu du XVIIIe siècle*, Paris : Institut National des Etudes Démographiques, 2011.

<sup>12</sup> La composition exacte de ce qu'on appelle aujourd'hui le groupe de Gournay reste encore assez floue. Dupont de Nemours (Schelle 1897: 238) y fait figurer Plumart de Dangeul, Herbert, Malesherbes, Boisgelin, Cicé, Maynon d'Invau, Trudaine de Montigny, Morellet, Price et Tucker. Cette liste est peu vraisemblable, Turgot n'y figure pas, Montaudouin de la Touche et Trudaine non plus. En revanche, Price et Tucker seraient plus aisément catalogués parmi les inspirateurs de Gournay que parmi ses disciples. Cicé et Boisgelin n'ont jamais publié d'ouvrages économiques, et malgré leur étroite relation avec Turgot, n'ont pas vraiment de raison d'apparaître ici. Schelle ajoute à cette liste l'abbé Coyer, Clicquot-Bervache et Montaudouin de la Touche. Morellet complète enfin cette liste des collaborateurs de Gournay en y ajoutant Véron de Forbonnais (Morellet 1988: 64).

Le premier groupe rassemble les hommes, qui comme Gournay, ont tenté de faire évoluer le système réglementaire de l'intérieur pour encourager l'industrie française. Leur position est le plus souvent modérée, et leur statut les oblige généralement à composer avec le pouvoir royal, et une forte opposition conservatrice. Cependant, ils ont tous joué un rôle important dans la diffusion et l'application des principes de Gournay sans pourtant en avoir été des théoriciens.

Les membres de la deuxième catégorie ont tous écrit ou traduit pour Gournay. Le choix opéré par Gournay de faire traduire des auteurs étrangers n'est pas anodin. On peut y voir deux raisons. La première est politique, il est toujours moins risqué de publier un ouvrage de politique économique étranger qu'un ouvrage français, car ce dernier pourrait être interdit. On ne peut en effet pas ouvertement critiquer l'action du gouvernement, car ce serait critiquer le choix du Roi. On peut en revanche suggérer ce qui pourrait être fait, ou expliquer ce que font les autres. Mais l'exercice est périlleux, témoin, la traduction de Child qui n'a pas eu de difficulté à paraître mais dont les *Remarques* rédigées par Gournay n'ont pas pu s'y joindre. La seconde raison qui peut être avancée sur le choix de la traduction comme instrument de diffusion de la pensée libérale, c'est l'insuffisante formation des auteurs français. La littérature économique anglaise est en 1750 incontestablement en avance sur la française, et bien qu'on ait des auteurs libéraux réputés avant 1750 comme Herbert, Boisguilbert, Melon, Dutot ou d'Argenson, leurs écrits se concentraient encore trop peu sur le problème du développement industriel et de la liberté de fabrication. Gournay s'entoure donc d'un certain nombre de ses amis à qui il demande de rassembler les meilleurs ouvrages étrangers sur le sujet et de les traduire. Il participe activement à cette activité malgré le peu de temps que lui laisse le bureau du commerce.

La traduction d'ouvrages économiques constitue une forme de vulgarisation, sous-estimée de l'histoire de la pensée, mais néanmoins centrale pour l'histoire intellectuelle. Celles-ci avaient le mérite de défendre la liberté de commerce sans se compromettre auprès du gouvernement et surtout du roi. Antoin Murphy a pu montrer que les traductions d'ouvrages économiques entre 1730 et 1750 faisaient l'objet d'une véritable stratégie de communication et de propagande organisée par Trudaine et exécutée par le groupe de Gournay<sup>13</sup>. Il semblerait que les traductions de Morellet aient également répondu à des demandes ou des motivations d'ordre politique<sup>14</sup>. Morellet a été un traducteur important à son époque, il a en effet réalisé pas moins de 42 traductions de son vivant (cinq de l'italien, six du latin, et vingt-neuf de l'anglais)<sup>15</sup>. Il est surtout connu pour avoir traduit *Des délits et des peines* de Beccaria, dont la version altérée a longtemps servi de référence en France, mais il est également le premier traducteur de *La Richesse des Nations* d'Adam Smith, qu'il n'a cependant publiée en totalité<sup>16</sup>, mais dont le chapitre sur les corporations fut traduit et publié avant même la publication anglaise de l'ouvrage de Smith, probablement sur sollicitation de Turgot<sup>17</sup>, alors encore contrôleur général des finances.

---

<sup>13</sup> A. Murphy, « Le développement des idées économiques en France (1750-1756) », *Revue d'Histoire Moderne et Contemporaine*, T.XXXIII, octobre-décembre, 1986, pp.521-541.

<sup>14</sup> D. Medlin, « André Morellet, translator of liberal thought », *Studies on Voltaire*, 1978, CLXXIV, pp.189-201.

<sup>15</sup> D. Medlin, « Catalogue of Morellet's works », in Medlin D., Merrick J., *André Morellet (1727-1819) in the Republic Letters and the French Revolution*, New-York : Peter Lang, 1995, pp.185-232. Voir aussi mon étude sur ses traductions italiennes, C. Salvat, « André Morellet et la diffusion de la pensée économique italienne », *Il pensiero economico italiano*, Pise, VIII, 2000, n°2, pp.99-124.

<sup>16</sup> La disgrâce de Turgot, et l'échec de ses réformes économiques, expliquent sans doute en partie le sort malheureux de sa traduction pourtant achevée ; C. Salvat, « Histoire de la traduction inédite de la Richesse des nations par l'abbé Morellet. 'Une traduction manuscrite toujours célébrée et toujours obstinément refusée au public.' », *Storia del Pensiero Economico*, Florence, 1999, n.38, pp.119-136.

<sup>17</sup> « J'ai fait toutes les diligences que vous avez désirées de moi pour hâter l'impression de l'extrait de Smith [;] mais l'administration de la librairie est une si belle chose qu'on ne peut pas imprimer une feuille unique sans essayer des longueurs de plusieurs semaines. [...] Je pense que cet extrait tel qu'il est mériterait que vous le fissions lire et produiroit un très bon effet sur l'esprit droit et bon

Bien que connaissant bien Vincent de Gournay, c'est avant tout Trudaine qui fut le principal commanditaire des travaux économiques de Morellet. Vincent de Gournay représente le fondateur du mouvement, le personnage central autour duquel les collaborateurs se rassemblent. C'est lui qui à l'origine des principes qui fonde ce groupe, mais ce n'est pas Gournay qui le dirige, c'est Trudaine. Si l'on osait se permettre la comparaison, nous dirions que Trudaine fut à Gournay ce que fut Dupont à Quesnay.

C'est Trudaine, tout d'abord, qui chargea Morellet d'écrire son premier ouvrage à caractère économique, les *Réflexions sur les avantages de la libre fabrication et de l'usage des toiles peintes en France*, publié en 1758<sup>18</sup>. Trudaine est un interlocuteur privilégié de Gournay, et ce dernier lui fait part de toutes ses pensées et de tous ses projets. Il ne se contente pas en effet de lui transmettre les mémoires qu'il récolte auprès des intendants de province ou des négociants, il lui envoie aussi ses propres travaux, y compris sa traduction de Child qu'il accompagne de ses notes. Son fils, Trudaine de Montigny (1733-1777) le seconda en tant qu'adjoint à l'intendance des finances, de 1757 à 1769, date à laquelle il le remplaça, avant d'être appelé par Turgot à son arrivée au contrôle général.

Deux exemples suffisent, je crois, à illustrer les visées politiques des écrits économiques de Morellet, le débat sur la dissolution du privilège exclusif de la Compagnie des Indes en 1769 et celui sur la libéralisation du commerce des blés en 1770-74. En 1769, le gouvernement, dirigé par un disciple de Gournay, Maynon d'Inveau, décide de supprimer le monopole de

---

du lecteur. Je prends cette occasion (et je crois que vous pourriez la prendre de votre côté auprès de qui il appartient) pour vous proposer de demander au roi deux mille ecus pour faire faire avec soin la traduction de l'ouvrage de Smith. » D. Medlin, J.C. David, et P.A. Leclerc, *Lettres d'André Morellet*, Oxford : Voltaire Foundation, vol. I, 1991, p.310.

<sup>18</sup> Le sujet avait déjà été abordé par Gournay dans un petit mémoire intitulé *Observations sur la liberté des toiles peintes*, et qui avait été publié à la suite de l'ouvrage de Véron de Forbonnais sur le même sujet, *l'Examen des avantages et des désavantages de la prohibition des toiles peintes*, en 1755.



commerce de la Compagnie des Indes, une fois de plus en difficulté de trésorerie<sup>19</sup>. Très affectée par la Guerre de Sept ans, la compagnie a à son passif une dette de 60 millions qu'elle est incapable de résorber par sa seule activité commerciale. Son commerce fortement déficitaire, l'entretien de ses comptoirs et la gestion de cette énorme administration rendent a priori impossible la rentabilité de cette entreprise. La compagnie souffre, par ailleurs, de son statut ambigu et du poids écrasant d'un Etat omniprésent dans sa gestion. Le contrôleur général fit alors appel à Morellet pour rédiger un ouvrage destiné à convaincre les actionnaires et l'opinion publique du bienfait de la suppression de la compagnie des Indes<sup>20</sup>. Le *Mémoire sur la situation actuelle de la compagnie des Indes* de Morellet, concluant à la suppression du privilège, fut envoyé à la compagnie le 20 juillet 1769, dont le privilège fut révoqué le 13 août suivant<sup>21</sup>.

La visée politique des ouvrages de Morellet est encore plus flagrante dans le cas de la libéralisation du commerce des grains, qui fut l'un des plus importants édits à être mis en place par Turgot à son arrivée au pouvoir. En janvier 1770, Ferdinando Galiani publie une féroce satire des principes 'universels' des économistes, et condamne toute application aveugle de la liberté de commerce. Avant même la parution des *Dialogues*, Morellet fut contacté par Trudaine de Montigny pour en rédiger une réfutation « pour seconder les vues raisonnables du ministère [*i.e.* Maynon d'Invaux], et en particulier de M. le duc de Choiseul,

---

<sup>19</sup> Sur l'histoire de la Compagnie des Indes, voir H. Weber, *La Compagnie française des Indes, 1604-1875*, Paris : Arthur Rousseau Editeur, 1904.

<sup>20</sup> C. Salvat, « L'affaire de la Compagnie des Indes », *Revue européenne* 2000, Décembre 2003, pp.5-6.

<sup>21</sup> Le privilège exclusif de la Compagnie fut à nouveau restauré en 1787 par le contrôleur général Calonne. Les députés des principales villes maritimes françaises, qui profitaient depuis une vingtaine d'années de la liberté de commerce au Levant, chargèrent André Morellet de défendre à nouveau leur cause. Ce dernier rédigea donc à nouveau deux mémoires sur le sujet, *Mémoires relatifs à la discussion du privilège de la nouvelle Compagnie des Indes* et *Réponse précise au précis pour les actionnaires de la nouvelle Compagnie des Indes*, tous deux publiés en 1787

en faveur de la liberté de commerce »<sup>22</sup>. Malheureusement pour lui, en décembre 1769, Maynon d'Invaux est remplacé par l'abbé Terray, qui est lui opposé à la libéralisation du commerce des blés. La *Réfutation* de Morellet, bien qu'imprimée et prête à être distribuée, fut interdite et embastillée.

A son arrivée au pouvoir en 1774, Turgot fit néanmoins ressortir les exemplaires de la *Réfutation* de Morellet de la Bastille pour les faire mettre en vente vers la mi-novembre. La re-publication de la *Réfutation de l'ouvrage intitulé « Dialogues sur le commerce des blés »* est annoncée le 20 décembre 1774 dans les *Mémoires Secrets* de Bachaumont, c'est-à-dire le lendemain de l'enregistrement du premier édit de Turgot sur la liberté de commerce des blés par le Parlement. Turgot envoya également quelques jours plus tard une circulaire à ses intendants leur demandant d'acheter chacun 40 exemplaires de la *Réfutation*. Un certain nombre d'exemplaires furent de plus envoyés aux chambres de commerce pour en assurer une large diffusion.

#### C. DEFINIR (ET IMPOSER) UNE DEFINITION DE L'ECONOMIE POLITIQUE

Les dictionnaires de commerce et encyclopédies constituent au dix-huitième siècle le troisième vecteur important de diffusion de la pensée économique<sup>23</sup>. Morellet présente son projet de dictionnaire comme une tentative de synthèse encyclopédique de toutes les connaissances pratiques et théoriques dont relève l'économie politique. Son ambition est

---

<sup>22</sup> A. Morellet, *Mémoires de l'abbé Morellet de l'Académie française sur le dix-huitième siècle et sur la Révolution*, éditées par J.P. Guicciardi, Paris : Mercure de France, 1988, p.170.

<sup>23</sup> J.C. Perrot, « Les dictionnaires de commerce au XVIIIe siècle », *Revue d'Histoire moderne et contemporaine*, 1, 1981, pp.36-67, republié dans J.C. Perrot, *Une histoire intellectuelle de l'économie politique*, Paris : Editions de l'Ecole des Hautes Etudes en Sciences Sociales, 1992.

aussi claire que démesurée. Non seulement, il souhaite prouver la scientificité de l'économie politique (à travers l'unicité de son discours et l'universalité de ses principes) mais il veut en donner un état aussi exhaustif que possible, quitte à l'améliorer au fur et à mesure des éditions successives. Le *Dictionnaire de commerce* que Morellet se propose de rédiger doit regrouper trois masses de connaissances liées à la notion générale de commerce : la *géographie commerçante*, les *objets et matières du commerce*, et la *théorie du commerce et de ses opérations*. Les deux premières catégories regroupent les *Faits* du commerce, la dernière s'occupe exclusivement de la *Théorie* du commerce.

La publication à venir du dictionnaire est officiellement annoncée par les Frères Etienne en juin 1763 dans le *Mercure de France*, et la première lettre dont on dispose y faisant allusion date d'avril 1765<sup>24</sup>. Mais l'entreprise lui a été réellement confié en 1762. Après avoir étudié le dictionnaire de Savary, et les corrections à y apporter, Morellet décida de refondre complètement l'ancien ouvrage qui selon lui était mal bâti, incomplet et au contenu parfois erroné<sup>25</sup>. Il lui reproche en particulier de ne pas s'intéresser assez à l'agriculture « cet objet le premier & le plus intéressant de ceux qui doivent entrer dans le tableau du commerce ; ils négligent le plus souvent de nous instruire de l'Administration du commerce, de la Jurisprudence commerçante, de la Navigation intérieure & extérieure, des Revenus publics, &c, de chaque pays. »<sup>26</sup>. Mais surtout, le plus grand défaut de ce dictionnaire selon Morellet, est de ne pas avoir développé une partie théorique dans laquelle seraient définies les nouvelles notions de l'économie politique.

---

<sup>24</sup> D. Medlin, J.C. David, et P.A. Leclerc, *Lettres d'André Morellet*, op.cit. pp.27-28.

<sup>25</sup> A. Morellet, *Mémoires de l'abbé Morellet de l'Académie française sur le dix-huitième siècle et sur la Révolution*, op.cit. p.163 ; A. Morellet, *Prospectus d'un nouveau dictionnaire de commerce*, Paris : Frères Estienne, 1769, pp.353-54.

<sup>26</sup> A. Morellet, *Prospectus d'un nouveau dictionnaire de commerce*, op. cit. p.16.

Le projet est finalement peu à peu abandonné<sup>27</sup>. Seul paraît en 1769 le *Prospectus d'un Nouveau dictionnaire de commerce*, accompagné d'un *Catalogue d'une bibliothèque d'économie politique*, comprenant 740 titres publiés entre 1530 et 1768. Ce catalogue lui permet de circonscrire le champ de l'économie aux seuls sujets afférents à la production et à la circulation des richesses. Cela se traduit, notamment, par l'exclusion des réflexions économiques à visée plus politiques, telles que celles de Rousseau.

L'économie politique ou économie publique est la « la science qui enseigne à acquérir[,] à conserver[,] à distribuer les richesses de la société toute entière comme l'économie privée enseigne à chaque particulier[,] à acquérir[,] à conserver et distribuer les richesses privées. »<sup>28</sup>.

L'objectif de l'Etat est assez clair et assez consensuel dans l'énonciation que Morellet en donne dans le *Prospectus* : « L'objet de l'économie politique, de l'Administration, ne peut être distingué du but que se sont proposé les individus en se réunissant en société, c'est-à-dire, en établissant une Administration quelconque. De là il résulte nécessairement que l'objet de l'Administration d'une société doit être l'acquisition & la conservation des richesses, comme c'est celui de chaque individu. »<sup>29</sup>. L'Etat n'a tout d'abord pas d'autre existence que celle des membres qui le constituent, et pas d'autre objectif que celui de satisfaire le bonheur de chaque individu. Morellet, conscient de l'impossibilité de satisfaire la totalité des besoins, est dès lors confronté au problème de l'agrégation et choisit la règle du maximum : « Le bonheur et le malheur des sociétés n'est que le bonheur et le malheur

---

<sup>27</sup> Nous ne savons pas exactement quand Morellet abandonna définitivement le dictionnaire, bien qu'il situe lui-même cette décision pendant la Révolution, A. Morellet, *Mémoires de l'abbé Morellet de l'Académie française sur le dix-huitième siècle et sur la Révolution*, op.cit., pp.162-63.

<sup>28</sup> Cité par C.Salvat, " 'De la science de l'économie publique' de Morellet ", *Studies on Voltaire and the Eighteenth Century*, October 2003, p.89..

<sup>29</sup> A. Morellet, *Prospectus d'un nouveau dictionnaire de commerce*, op. cit. p.203.

des individus qui la composent, et non de cet être abstrait qu'on appelle société politique, nation. Le bonheur et le malheur d'une société sont le bonheur ou le malheur du plus grand nombre des citoyens. »<sup>30</sup>.

## 2. L'ÉCONOMIE POLITIQUE DES LUMIÈRES OU LE COMBAT DES ENCYCLOPÉDISTES

La représentation que l'on se fait aujourd'hui de l'économie politique du XVIII<sup>e</sup> a été fortement influencée par les économistes du XIX<sup>e</sup> siècle. Les physiocrates furent ainsi, pendant très longtemps, les seuls auteurs du dix-huitième siècle étudiés pour leur contribution à l'économie politique, et cela bien que l'importance théorique du Tableau économique n'ait été reconnue que très tard, ainsi que nous l'avons montré dans le cas de Walras<sup>31</sup>. La délimitation étroite de l'économie politique, imposée au début du siècle par Jean-Baptiste Say, a amené les premiers historiens de la pensée économique à restreindre de manière excessive le champ de leur investigation, et à se limiter à une histoire des concepts économiques. Certaines notions se sont ainsi avérées plus aisées à manipuler par les économistes car elles faisaient écho à des concepts bien connus tels que le circuit économique de Quesnay, l'équilibre de Turgot, la valeur utilité de Condillac etc. La réflexion épistémologique a été également largement sous-traitée. Mon travail de thèse a consisté à montrer que l'histoire de la pensée économique ne devait pas se limiter à l'étude des 'grands' auteurs ni des seuls écrits à caractère théorique.

---

<sup>30</sup> Cité par C.Salvat, " 'De la science de l'économie publique' de Morellet ", *op.cit.*, p.91.

<sup>31</sup> C. Salvat, "L'ambivalence de la référence physiocratique chez Walras", in *Études walrassiennes*. Roberto Baranzini, Arnaud Diemer et Claude Mouchot (éds.), Paris: L'Harmattan, 2004, pp. 29-50. Voir aussi C. Salvat, « 'On ne fait point d'industrie entre ciel et terre' : Walras, lecteur de Quesnay », *Cahiers du CERAS*, Hors série n°3, Mai 2004, pp.207-221.

Après mon admission au CNRS, j'ai souhaité aller plus loin dans ce que je considérais alors comme étant la meilleure approche épistémologique possible, à savoir l'histoire intellectuelle de l'économie politique. Lors de mes précédents travaux, j'avais montré que Morellet fut le premier à donner une définition (et par conséquent une existence officielle) à l'économie politique en tant que science, science dont il limitait alors strictement l'objet à l'acquisition et à la conservation des richesses<sup>32</sup>. Ironiquement, il fut donc le premier à adopter la perspective même que mes travaux avaient précisément pour objet de déconstruire. L'un des principaux enseignements de l'histoire intellectuelle est, en effet, que pour comprendre ce que fut l'économie politique à une époque donnée, on ne doit pas se limiter au seul corpus établi par les économistes eux-mêmes mais que l'on doit en rechercher partout les expressions.

Le but de ce projet de recherche a alors été de comprendre comment un ensemble de savoirs pratiques et théoriques, incluant tous les savoirs relatifs au commerce, à la finance, à l'industrie, à l'agriculture mais également à l'administration domestique et à l'administration publique, s'est soudainement organisé et hiérarchisé en tant que science, et comment ce que l'on a parfois appelé l'autonomisation du savoir s'est traduite par la dépréciation et la mise à l'écart de certaines catégories de savoir au profit d'autres plus en adéquation avec le nouveau statut de ces connaissances.

Pourquoi l'histoire de la pensée économique française débute-t-elle par François Quesnay plutôt que, par exemple, Véron de Forbonnais ou Rousseau ? La qualité théorique des travaux seule ne permet pas de le justifier. Leur impact non plus. L'explication la plus

---

<sup>32</sup> C. Salvat, " 'De la science de l'économie publique' de Morellet ", *Studies on Voltaire and the Eighteenth Century*, October 2003, pp.65-96.

plausible est qu'aucun auteur n'est pas à proprement parler responsable de 'naissance' de l'économie politique, mais que la reconnaissance collective de ce savoir en tant que science, que l'on peut plus ou moins préciser dater, a constitué une ligne de démarcation entre les premiers économistes et les derniers précurseurs. La véritable question n'est donc pas de savoir quel fut le premier économiste mais pourquoi et comment des savoirs déjà existants ont, à un moment donné, été identifiés comme appartenant à la science de l'économie politique, et pourquoi d'autres en ont été écartés.

Dans ce qui suit, je présente deux visions alternatives de l'économie politique, développées toutes deux dans l'*Encyclopédie* de Diderot et D'Alembert dans les années 1750. Elles incarnent deux visions politiques et deux philosophies des sciences différentes. La première fait de l'économie politique une science sinon 'exacte' du moins rigoureusement fondée dans les faits, c'est celle de Quesnay, et l'autre en fait un savoir relatif à l'organisation politique, c'est celle de Rousseau. La première s'est très vite imposée, notamment grâce aux efforts de penseurs tels que Morellet et à ses soutiens politiques et institutionnels.

a. L'ECONOMIE POLITIQUE, NOUVELLE BRANCHE DES SAVOIRS DANS L'ENCYCLOPEDIE DE DIDEROT ET D'ALEMBERT

Mes recherches sur Morellet, et notamment sur son projet de dictionnaire de commerce, m'ont naturellement conduit à m'interroger sur la question de la définition de l'économie politique avant 1769. Or, les deux décennies qui ont précédé la publication du *Prospectus* ont été marquées par la publication d'une des plus vastes entreprises philosophiques de l'époque moderne, véritable catalogue raisonné de toutes les connaissances, à savoir l'*Encyclopédie* de Diderot et d'Alembert.

L'hypothèse que j'ai voulue explorer alors est que la réorganisation des savoirs opérée à l'occasion de l'édition de l'*Encyclopédie* de Diderot et d'Alembert entre 1751 et 1759<sup>33</sup> fut à l'origine en France de la reconnaissance économie politique en tant que science. Quesnay en a certes profité, puisque ses articles *Fermiers* et *Grains* ont été classés comme appartenant à l'économie politique, mais il n'en a pas été lui responsable. L'étude que j'ai menée<sup>34</sup> montre, en effet, que les désignants (autrement dit le référencement des articles au sein d'un classement des sciences) étaient attribués par les éditeurs et non par les auteurs des articles. Dans le cas de Quesnay, l'existence de deux articles *Fermier(s)*, l'un au singulier signé par Leroy et l'autre au pluriel signé de Quesnay, explique probablement le classement du second en 'économie politique' alors que le premier fut logiquement référencé sous 'économie rustique'. Par souci de cohérence, probablement, le même désignant fut ensuite attribué à l'article *Grains*. Les autres articles rédigés par Quesnay ('Hommes'..), et que leur auteur retira suite à la suite à la révocation du privilège, ne possèdent aucun désignant. L'expression 'économie politique' n'est, par ailleurs, jamais employée dans le corps de texte de ses articles.

Pourquoi certaines connaissances furent-elles considérées comme appartenant à l'économie politique (un champ disciplinaire qui n'existait jusqu'alors pas) et pourquoi certaines autres furent-elles été rangées du côté de l'économie rustique, du commerce ou de l'économie domestique? Telle a été la question que je suis posée en me lançant dans cette longue et fastidieuse étude. Il faut dire, tout d'abord, que le corpus était énorme.

---

<sup>33</sup> Soit lors de la publication des sept premiers volumes, le privilège de l'*Encyclopédie* ayant été révoqué en 1759 suite à l'attentat de Damiens (1757) et la publication de *De l'esprit* d'Hévétius. Sur l'histoire de l'*Encyclopédie* voir R. Darnton, *L'aventure de l'Encyclopédie, 1775-1800: un best-seller au siècle des Lumières*. Translated by M.A. Revellat: Librairie académique Perrin, 1992.

<sup>34</sup> C. Salvat, "Oeconomies. Les articles OE/Economie et leurs désignants dans l'*Encyclopédie*", *Recherches sur Diderot et l'Encyclopédie*, n°40, Octobre 2006, pp.97-116.



L'*Encyclopédie* s'étend sur 17 volumes de textes et 11 volumes de planches, publiés entre 1751 et 1782. Elle comprend 72000 articles écrits par plus de 140 auteurs, dont bien sûr Diderot et d'Alembert, mais également Voltaire, Rousseau, Montesquieu, Quesnay, Forbonnais... Elle représente presque à elle seule la pensée des Lumières dans toute sa richesse et sa diversité. De nombreux chercheurs, de toutes les disciplines, se sont spécialisés sur ce seul ouvrage. L'économie fut l'une des rares disciplines à ne pas en avoir fait l'objet d'une étude exhaustive<sup>35</sup>.

Le travail que j'ai alors entrepris sur ce corpus a consisté à faire un inventaire aussi exhaustif que possible des articles considérés comme économiques à l'époque. Ceux-ci incluent, outre les 'grands' articles déjà cités, la multitude de petits articles sur le commerce, l'industrie, l'agriculture, et de s'interroger sur leur relation respective au genre 'économie', celui-ci pouvant se subdiviser en économie publique ou politique, économie rustique ou économie domestique, les uns renvoyant généralement aux autres. Dix articles ont pour désignant 'économie politique'. Il s'agit de FINANCES de Pesselier, INTÉRÊT, MENDIANT, RENTIER, SEL de Jaucourt, FERMIERS et GRAINS de Quesnay et VINGTIÈME signé Boulanger mais en partie de Diderot. À ceux-ci on peut rajouter deux petits articles à désignants composés, RASPHUIS (*Hist. mod. Economic politiq.*) et SPINHUYS (*Hist. mod. Econom. politiq.*). FINANCES est lui-même désignant d'une cinquantaine d'articles. Les articles de Véron de Forbonnais ont pour la plupart tous été publiés dans les cinq premiers volumes de l'*Encyclopédie*, soit avant le premier emploi du désignant économie politique pour l'article FERMIERS de Quesnay. Ils sont néanmoins liés à différents niveaux aux articles

---

<sup>35</sup> On notera cependant l'exception que constituent les travaux de la linguiste Marie-France Piguet. M.F. Piguet, « OEconomie/Économie (politique) dans le texte informatisé de l'*Encyclopédie* », *Recherches sur Diderot et sur l'Encyclopédie*, p. 31-32, Avril 2002, pp. 123-137.

'économiques'. Les articles ÉCONOMIE RUSTIQUE et FERMIER (*économie rustique*), renvoient ainsi à CULTURE DES TERRES (*commerce politique*) qui renvoie lui-même à AGRICULTURE de Diderot et à GRAINS (*économie politique*) de Quesnay. L'article COMMERCE également de Forbonnais renvoie également à AGRICULTURE, qui renvoie lui-même à différents articles d'économie rustique (FEUILLE, FROMENT, SEMOIR) ainsi qu'à l'article GRAIN (sic). On observe de la part des éditeurs de l'Encyclopédie une volonté de ne pas employer le seul terme d'oeconomie (une dizaine d'occurrences sur trois cents) au profit des désignants composés économie rustique et économie domestique. L'article FORET de Leroy constitue l'une des rares exceptions. Les désignants économie rustique, parfois abrégés par économie (ex. FARINE, GRANGE) renvoient généralement à un ensemble de techniques productives de l'agriculture, parfois fondées sur les dernières recherches scientifiques, et non au savoir familial et ménager de l'économie domestique. Enfin, les désignants 'économie domestique' sont peu nombreux, une trentaine, et marquent le plus souvent des articles ménagers (CORBEILLE, CRUCHE, ÉTEIGNOIR, GOBELET, MOUCHOIR...). L'économie domestique, dans l'emploi qui en est fait dans la désignation des articles, est – on le voit – assez lointaine de l'approche aristotélicienne de Rousseau. Les articles d'économie rustique, le plus souvent rédigés par Diderot, représentent les trois quarts des articles dits d'économie dans l'*Encyclopédie*. On en compte environ 220, dont AGNEAU, CHASSE, ENGRAIS, FERMIER, LABOUR, METAYER, SEMAILLE ou TROUPEAUX.

L'économie fait, également, l'objet de deux entrées 'en propre' dans l'*Encyclopédie*, usant ainsi de la polygraphie du terme, pour compléter ou corriger dans le tome XI ce qui avait été publié, dix ans plus tôt, dans le tome V. La première entrée se compose du célèbre article ÉCONOMIE (*Morale et Politique*) de Rousseau, réédité en 1758 sous le titre *Discours*

sur l'économie politique<sup>36</sup>, et d'une sous-entrée ÉCONOMIE RUSTIQUE. La seconde entrée du mot, à la lettre O, répond davantage au souhait exprimé par Diderot dans l'article ENCYCLOPÉDIE de mieux intégrer les termes de la langue usuelle. Elle se décline en quatre parties, la définition générale, et trois sous-entrées, OECONOMIE (*Critiq. sacrée*) de Jaucourt, OECONOMIE ANIMALE (*Médec.*) de Ménuret de Chambaud, et OECONOMIE POLITIQUE (*Hist. Pol. Rel. ant. & mod.*) de 'feu Boulanger'.

Pourquoi deux articles sur l'économie politique ? La polygraphie du terme œ/économie au dix-huitième n'est, bien sûr, qu'un prétexte pour Diderot. On a pu penser que celui-ci était mécontent des thèses développées par Rousseau, et notamment de son interprétation de la volonté générale. Plus fondamentalement, je crois qu'entre 1755, année de publication du tome V, et 1765, année de publication du tome XI, l'économie politique a pris une nouvelle signification, qu'il était essentiel pour l'*Encyclopédie* de reconnaître. L'article de Rousseau traite de l'administration de la société politique. Il prend donc le mot économie dans son sens premier, et l'applique à la nation, par opposition à l'économie domestique qui ne s'applique qu'à la famille. Il est d'ailleurs important de noter que Rousseau n'emploie pas l'expression d'économie politique mais d'économie *publique*, qui fait

---

<sup>36</sup> On doit en fait à Jacob Vernes, et à son édition de 1758 du *Discours sur l'économie politique*, la confusion entre l'article ÉCONOMIE et l'expression d'économie politique. Rousseau ne semble pas cependant avoir été associé au choix du titre.

davantage pendant à l'économie *domestique*<sup>37</sup>. Il suit, en cela, les dictionnaires de Furetière<sup>38</sup> et de Trévoux<sup>39</sup> dont l'*Encyclopédie* s'est beaucoup inspiré<sup>40</sup>.

Le second article, l'article OECONOMIE POLITIQUE fut publié en 1765, soit six ans après la mort de son signataire posthume, Nicolas-Antoine Boulanger. Il n'y a cependant aucun doute sur son attribution. L'article est une refonte, composée de larges extraits, du manuscrit des *Recherches sur l'origine du despotisme oriental* qui sera également publié à titre posthume en 1761. L'intitulé n'est cependant pas de Boulanger qui n'utilise jamais l'expression 'œconomie politique' dans le *Despotisme oriental*. Le corps du texte a peu de rapport avec l'intitulé de l'article, dont le véritable intérêt est tout entier dans sa définition, probablement de Diderot lui-même: «OECONOMIE POLITIQUE (*Hist. Pol. Rel. anc. & mod.*) c'est l'art & la science de maintenir les hommes en société, & de les y rendre heureux, objet sublime, le plus utile & le plus intéressant qu'il y ait pour le genre humain »<sup>41</sup>. Cette définition ne fut pas retenue par les principaux théoriciens de l'économie, dont Morellet, qui préférèrent la circonscrire à la seule question des richesses. Elle atteste néanmoins,

---

<sup>37</sup> L'article de Rousseau 'ÉCONOMIE ou OECONOMIE (Morale & Politique)' a souvent été abusivement désigné par ÉCONOMIE POLITIQUE, et ce à l'intérieur même de l'*Encyclopédie*. En réalité, Rousseau n'emploie jamais l'expression d'économie politique dans cet article, mais celui d'économie publique, qu'il utilise à sept reprises. Une hésitation, cependant, au début de son article : « économie générale, ou politique », à laquelle Rousseau ne donnera pas suite.

<sup>38</sup> L'édition de 1690 du *Dictionnaire universel de Furetière* définit ainsi le substantif 'œconomie' comme suit : « OECONOMIE. Subst. Masc. [sic ?] Ménagement prudent qu'on fait de son bien, ou de celui d'autrui. L'œconomie est la seconde partie de la Morale, qui enseigne à bien gouverner une famille, une Communauté. Ce prier entend bien l'œconomie. Quelquefois on couvre l'avarice du nom honneste d'œconomie. » *Dictionnaire universel contenant tous les mots françois tant vieux que modernes, & les termes de toutes les sciences et des arts* [par A. Furetière], [1690], tome I, [non paginé].

<sup>39</sup> Le *Dictionnaire de Trévoux* fut définit, dans son édition de 1752, l'économie comme la « conduite sage, [le] ménagement prudent qu'on fait de son bien ou de celui d'autrui.[...] L'économie est la seconde partie de la Morale, qui enseigne à bien gouverner une famille, une communauté », *Dictionnaire universel françois et latin* [dit dictionnaire de Trévoux], 5e éd., 1752, III, 494.

<sup>40</sup> M. Leca-Tsiomis, *Ecrire l'Encyclopédie. Diderot : de l'usage des dictionnaires à la grammaire philosophique*, Oxford, Voltaire Foundation, 1999.

<sup>41</sup> OECONOMIE POLITIQUE, XI, 366-67

déjà, de l'existence d'une conception alternative de l'économie politique, qui portait sur les modalités et finalités économiques de la vie en société (comment rendre les gens heureux) plutôt que sur la légitimité de l'association civile. C'est vers cette seconde conception, presque immédiatement abandonnée, que je me tourne maintenant.

b. LA CONCEPTION ALTERNATIVE DE L'ÉCONOMIE POLITIQUE CHEZ JEAN-JACQUES ROUSSEAU

Dans son article, Rousseau ne se propose pas d'interroger le principe, ni les différentes formes de la société politique – comme c'est le cas dans l'article de Boulanger –, mais son administration. Comment une société formée peut-elle remplir de manière pérenne son objet ? L'économie, c'est-à-dire selon la définition de Trévoux « l'ordre par lequel un corps politique subsiste principalement », permet de répondre à la question. Elle se compose, chez Aristote comme chez Rousseau, de deux rapports, le rapport aux personnes et le rapport aux biens. En ce sens, l'économie est synonyme de gouvernement. L'économie est, écrit Rousseau, « le sage et légitime gouvernement de la maison »<sup>42</sup> s'inspirant probablement lui aussi de Trévoux (le « sage et prudent gouvernement de la famille », *prudens administratio*). Le terme d'économie, poursuit Rousseau, « a été dans la suite étendu au gouvernement de la grande famille, qui est l'état. Pour distinguer ces deux acceptions, on l'appelle dans ce dernier cas, économie générale, ou politique ; et dans l'autre cas, économie domestique, ou particulière »<sup>43</sup>. La substitution de la prudence par la légitimité du gouvernement est un point notable de la pensée de Rousseau. La dimension normative de l'économie se trouve toute entière concentrée dans la question politique de

---

<sup>42</sup> ÉCONOMIE, *Enc.*, V, 337a.

<sup>43</sup> ÉCONOMIE, *Enc.*, V, 337a.

sa justice et de sa légitimité. C'est à ce niveau, nous semble-t-il, que l'on doit interroger le rapport de l'économie domestique à l'économie politique, autrement dit le statut du modèle de famille dans l'administration publique. L'économie d'une famille s'assure spontanément par l'autorité naturelle exercée par le chef de famille, et l'intérêt commun qui unit les membres de la famille. En revanche, une société politique se caractérise par la convention et la divergence d'intérêt entre gouvernants et gouvernés. D'un côté la subsistance du corps constitue l'objectif commun des membres, de l'autre il se fait à leur détriment : « Le principal objet des travaux de toute la maison, est de conserver et d'accroître le patrimoine du pere, afin qu'il puisse un jour le partager entre ses enfans sans les appauvrir ; au lieu que la richesse du fisc n'est qu'un moyen, souvent fort mal entendu, pour maintenir les particuliers dans la paix et dans l'abondance »<sup>44</sup>. Enfin, la pérennité se pose à des niveaux très différents pour l'une ou l'autre, « la petite famille est destinée à s'éteindre, et à se résoudre un jour en plusieurs autres familles semblables ; mais la grande [est] faite pour durer toujours dans le même état »<sup>45</sup>. Le gouvernement de la famille est naturel, il s'appuie sur l'inégalité physique et morale du chef. Il est par nature légitime.

L'objet de la société politique doit donc d'être de retrouver, par l'artifice, l'autorité et la légitimité naturelles du père de famille. Mais, pour gouverner l'État comme une famille, il faut que la nation devienne elle-même une 'grande famille', qu'elle transcende l'artificialité du lien social par la vertu civile. Rousseau distingue une économie publique légitime, pensée à l'image de l'économie domestique, d'une économie publique trompeusement inspirée du modèle domestique, celle du paternalisme politique<sup>46</sup>. L'idéal de l'économie

---

<sup>44</sup> ÉCONOMIE, *Enc.*, V, 337b.

<sup>45</sup> ÉCONOMIE, *Enc.*, V, 337b.

<sup>46</sup> De manière analogue, pour Boulanger, les lois civiles établies à l'image des lois domestiques ne permettent pas de répondre aux spécificités des sociétés politiques. Parce que les hommes ont fait de la coutume et de l'histoire une légitimité en soi, au détriment de la raison même, l'âge d'or s'est

politique entretient donc un rapport ambivalent avec l'économie domestique avec laquelle elle avait d'abord été opposée. La légitimité de l'économie publique ou politique est fondée sur la vertu, soit la conformité des volontés particulières à la volonté générale. Elle donne à l'autorité politique une force comparable à l'autorité naturelle du père qui jouit d'une obéissance totale au sein de la famille sans avoir recours à la contrainte ou à la menace. La vertu d'un peuple, autrement dit sa citoyenneté, peut être développée par deux moyens : l'éducation publique et le patriotisme. La patrie constitue le cadre de référence dans lequel une communauté d'individus, tous mus par leur seul intérêt particulier, peut atteindre l'idéal politique que représente l'intérêt général. En cela, elle se substitue à la famille. Toute la question de la légitimité et de la pérennité de la société se réduit donc à celle-ci : comment former des citoyens ? L'éducation des enfants est abandonnée « aux lumières et aux préjugés des pères » alors qu'elle « importe à l'état encore plus qu'aux pères »<sup>47</sup>. La mort du père interrompt de plus cette formation à un moment crucial de l'éducation civique, « l'état demeure, et la famille se dissout »<sup>48</sup>. Rousseau propose donc de transférer cette fonction paternelle à l'État et de substituer l'éducation publique à l'éducation domestique. Ce faisant, il transfère l'autorité paternelle à l'autorité politique. Désormais, le père se fera obéir de ses enfants par l'intermédiaire de

---

progressivement et insensiblement transformé en âge de plomb : « À mesure que le séjour de l'homme s'est embelli, à mesure que les sociétés se sont multipliées, & qu'elles ont formé des villes & des états, le règne moral a dû nécessairement faire place au règne politique, & le tien & le mien ont dû paroître dans le monde, non d'abord d'homme à homme, mais de famille à famille & de société à société, parce qu'ils y sont devenus indispensables, & qu'ils font partie de cette même harmonie qui a dû rentrer parmi les nations renouvelées, comme elle est insensiblement rentrée dans la nature après le dernier chaos. Cet âge d'or a donc été un état de sainteté, un état surnaturel digne de notre envie, & qui a justement mérité tous les regrets de l'antiquité : cependant lorsque les législations postérieures en ont voulu adopter les usages & les principes sans discernement, le bien s'est nécessairement changé en mal, & l'or en plomb », *OÉCONOMIE POLITIQUE, Enc.*, XI, 368a.

<sup>47</sup> *ÉCONOMIE, Enc.*, V, 343b.

<sup>48</sup> *ÉCONOMIE, Enc.*, V, 343b.

la loi L'administration publique, la troisième règle de l'économie publique, peut alors légitimement s'inspirer de l'économie domestique : « Ainsi après avoir parlé de l'économie générale par rapport au gouvernement des personnes, il nous reste à la considérer par rapport à l'administration des biens »<sup>49</sup>. L'économie publique a pour objet essentiel l'impôt et son mode de perception<sup>50</sup>. Celle-ci est au coeur de la question économique, c'est-à-dire de la gestion prudente des revenus et des dépenses, mais également de la nature du contrat social. Le dilemme qui se pose à Rousseau est donc le suivant : comment un État peut-il subvenir à ses besoins sans atteindre à la propriété privée de ses membres ? Comment assurer le bon fonctionnement de l'État en taxant le moins possible et le plus justement possible ? Par « une juste et sage économie »<sup>51</sup>. La nation, devenue 'grande famille' doit être gérée comme telle, c'est-à-dire avec économie, en bon père de famille<sup>52</sup>. Rousseau fait de l'économie domestique, au sens restreint du terme, i.e. la gestion du patrimoine familial, le modèle de l'administration politique. La nation peut ainsi, comme la famille, se réunir autour d'un patrimoine commun. Idéalement, l'État devrait tirer ses revenus d'un domaine foncier public dont la gestion constitue aux yeux de Rousseau « le vrai système économique »<sup>53</sup>. L'économie, ici entendue dans son acception moderne, est donc l'un des vecteurs du gouvernement légitime. En faisant de l'administration

---

<sup>49</sup> ÉCONOMIE, *Enc.*, V, 344a.

<sup>50</sup> Il existe alors deux systèmes de perception alternatifs, la régie et la ferme défendues respectivement par Rousseau et Quesnay. Dans les articles 'Fermes du Roi' (*Finances*) et 'Financier' (*Politiq.*), Pesselier critique sévèrement Montesquieu pour ses attaques contre la ferme et les financiers. Il reprend, dans ce premier article, un à un les huit principes du philosophe bordelais défendant la régie, dont le premier : « La régie est l'administration d'un bon père de famille, qui leve lui-même avec économie & ordre ses revenus » (FERMES, *Enc.*, VI, 513b). Un contributeur inconnu lui répondra dans l'article SUBSIDES.

<sup>51</sup> « si les contributions sont volontaires, elles ne produisent rien ; si elles sont forcées, elles sont illégitimes ; et c'est dans cette cruelle alternative de laisser périr l'état ou d'attaquer le droit sacré de propriété, qui en est le soutien, que consiste la difficulté d'une juste et sage économie », ÉCONOMIE, *Enc.*, V, 344d.

<sup>52</sup> ÉCONOMIE, *Enc.*, V, 345a.

<sup>53</sup> ÉCONOMIE, *Enc.*, V, 346a.



domestique un principe politique, l'économie, entendue comme prudens administratio, contribue à la citoyenneté de ses membres et à l'autorité de la loi. Le système commercial, fondé sur une division poussée du travail, une circulation accrue des marchandises et de la monnaie, est en grande partie responsable, selon Rousseau, de la perversion des mœurs (vanité, amour-propre) et de la décadence des civilisations occidentales. La société politique se réunit autour de la loi, alors qu'elle se déchire autour des richesses<sup>54</sup>.

### 3. L'ÉCONOMIE ET LA LOI CHEZ ROUSSEAU

J'ai débuté en 2006 un programme de recherche sur la pensée économique de Rousseau et son articulation avec son système philosophique et politique. Ce programme initié par des travaux sur la représentation de l'économie politique dans l'*Encyclopédie* de Diderot et D'Alembert m'a conduit à analyser les relations entre les réflexions, apparemment assez frustrées de Rousseau en économie, et ses théories politiques, beaucoup plus fines et développées. Un certain nombre d'auteurs défendent l'idée que les analyses de Rousseau, centrées sur l'autarcie, le rôle pernicieux de la monnaie et des échanges en général, devaient être réhabilités au regard de son système philosophique. La tentative est grande, disent-ils, de juger ces travaux à l'aune du discours économique actuel, lui-même produit par un système économique, le capitalisme, que Rousseau rejetait. Derrière ce qui peut

---

<sup>54</sup> « La puissance civile s'exerce de deux manières : l'une légitime par l'autorité, l'autre abusive par les richesses. Partout où les richesses dominant, la puissance et l'autorité sont ordinairement séparées, parce que les moyens d'acquérir la richesse et les moyens de parvenir à l'autorité n'étant pas les mêmes sont rarement employés par les mêmes gens. Alors que la puissance apparente est dans les mains des magistrats et la puissance réelle est dans celle des riches. Dans un tel gouvernement tout marche au gré des passions des hommes, rien ne tend au but de l'institution » J.-J. Rousseau, *Projet de Constitution pour la Corse, Oeuvres Complètes*, Paris, Gallimard, 1964, III, 939.

apparaître comme une bataille d'érudits discutant *ad nauseam* des interprétations possibles d'un corpus finalement assez limité (les œuvres complètes de Rousseau sont regroupées en cinq volumes de la Pléiade), se déroule en fait un débat beaucoup plus important, celui du relativisme scientifique appliqué à l'économie. Si l'on admet que Rousseau (cela pourrait, en l'occurrence, être n'importe quel auteur) a construit un système économique cohérent qui échappe totalement aux standards de la science économique contemporaine et que ce système ne prend son sens que dans un système politique distinct, voire antinomique, au système capitaliste qui définit nos sociétés actuelles, alors on est en droit d'admettre, ce que de nombreux sociologues et philosophes ont théorisé dans les années 70 et 80, à savoir que toute connaissance est le produit de pouvoirs institutionnels. Autrement dit, que la science économique aurait pour vocation première d'asseoir et de renforcer les pouvoirs de ceux qui elle profite. En tant qu'historien de la pensée économique, formé au Centre Léon et Walras de Lyon, et très orienté vers l'histoire contextuelle ou intellectuelle de la pensée, j'ai moi-même longtemps été, consciemment ou non, attiré vers ce cadre de pensée. Mes travaux de thèse sur la formation et la diffusion de la pensée économique libérale au dix-huitième siècle en portent la trace. Mon programme de recherche sur Rousseau était, d'une certaine manière, une façon pour moi de tester cette hypothèse 'de l'autre côté' de la barrière, celui des opposants de l'économie libérale alors défendue par les premiers économistes. La tentation a été grande, chez un certain nombre d'historiens de la pensée économique, de voir en Rousseau un 'autre' économiste, une autre façon de théoriser l'économie, et de comparer ses mérites à ceux de l'économie dite dominante ou orthodoxe. Les résultats de mes propres recherches sont autres, et montrent la profonde méconnaissance de Rousseau des mécanismes économiques. Rien ne permet non plus d'affirmer que le

système d'autoproduction qu'il décrit dans la *Nouvelle Héloïse* ait la dimension politique que certains veulent lui prêter. On a donc tort de vouloir faire de Rousseau un économiste oublié, première victime d'un système économique qu'il tenait de combattre, et toute reconstruction de son discours fondée sur les bribes de textes de natures différentes (traités, fictions, correspondances etc.) pose, de mon point de vue, de vrais problèmes méthodologiques. Ces conclusions ne sont pas celles que je comptais formuler au début des mes recherches. Elles se sont progressivement imposées à moi. J'ai parfois eu du mal à les accepter, mais je dois dire que j'ai encore eu plus de mal à les faire accepter à mes collègues historiens de la pensée économique, qui y voyaient non seulement une remise en cause de leurs propres travaux mais également de leur idéologie. Cette expérience a été certes difficile, notamment par la difficulté à publier par laquelle elle s'est traduite. Elle m'a néanmoins permis de m'éloigner du relativisme scientifique que les SHS, et notamment l'histoire de la pensée économique, contribuent à développer, et dont on commence aujourd'hui à mesurer les dangers.

a. L'ECONOMIE ET LA LOI CHEZ ROUSSEAU

La pensée économique de Rousseau a souvent été réduite à quelques fragments sur la monnaie, les inégalités, le luxe et l'autarcie. Comparée aux écrits contemporains de Quesnay, Smith ou Hume, elle est donc apparue comme 'retardataire' voire réactionnaire. Jean Mathiot a cependant montré que loin de s'inscrire dans une tradition passéiste, la pensée économique de Rousseau marque au contraire une avancée significative de la philosophie politique dans ses rapports à l'économie politique, dont elle accompagne, voire

anticipe, le développement<sup>55</sup>. Selon lui, la modernité de la pensée économique de Rousseau se révèle paradoxalement dans son ouvrage le plus politiquement achevé, et économiquement le plus épuré, le *Contrat Social*. Notre propre travail s'inscrit, sur ce point, dans sa continuité<sup>56</sup>. Nous nous en démarquons cependant sur un point, le statut de la rationalité économique. L'ambition de cet article est, en effet, de montrer, à travers le cas de l'échange, que la relation entre la politique et l'économie, bien que hiérarchisée, ne se réduit pas à une hégémonie de la rationalité politique. La rationalité économique apparaît, au contraire, constitutive de l'état civil légitime y compris dans son expression la plus négative, à savoir l'autarcie<sup>57</sup>.

La valorisation sociale et individuelle des richesses, loin d'être source de progrès, contrarie le développement normal de l'économie nationale. Rousseau se démarque donc sensiblement de ses contemporains tels que Smith ou Hume. La richesse et la loi constituent deux référents exclusifs l'un à l'autre<sup>58</sup>. Il n'est donc pas possible d'avoir dans le même temps deux sources d'autorité, la loi et la richesse. Or, seule la première est

---

<sup>55</sup> J. Mathiot, « Politique et économie chez Jean-Jacques Rousseau », dans J. Boulad-Ayoub, I. Schulte-Jenckhoff et P-M. Vernes (dir.), *Rousseau. Anticipateur-retardataire*, Paris, L'Harmattan, 2000, pp. 19-39.

<sup>56</sup> C. Salvat, « La loi et l'échange: le statut de la rationalité économique chez Rousseau », *Revue économique*, volume 58, n°2, Mars 2007, pp.381-398.

<sup>57</sup> Dans l'article L'échange et la loi, je montre en particulier, que Les effets pervers de l'échange, notamment le développement des inégalités morales, ne sont pas le fait de l'échange, qui dans son principe ne fait que reproduire les inégalités naturelles, mais de son instrumentalisation sociale. L'échange est, au même titre que la monnaie, les sciences et les arts, voire que le langage même, un outil de distinction sociale. L'emprise des passions (envie, vanité, orgueil) sur la raison trompe les individus sur leur propre intérêt (amour de soi) et dénature l'ensemble des institutions civiles. L'échange, en tant que mécanisme économique, est altéré et ne répond plus que partiellement à sa rationalité originelle. L'institution d'un cadre politique légitime pourrait, en théorie, atténuer l'impact des inégalités morales sur les représentations symboliques véhiculées par la richesse et la monnaie.

<sup>58</sup> « Jamais, écrit Rousseau dans la Lettre à d'Alembert, dans une monarchie l'opulence d'un particulier ne peut le mettre au-dessus du prince ; mais dans une République elle peut aisément le mettre au-dessus des lois. Alors le gouvernement n'a plus de force, et le riche est toujours le vrai souverain. » (Rousseau [1967], p. 171.)

légitime<sup>1</sup>. Quelle est alors la place de l'échange dans l'état civil ? Nous montrerons que la loi assure, en quelque sorte, la rationalité de l'échange en limitant la pratique aux usages conformes à l'amour de soi. La seconde partie de cette section s'interroge sur le statut paradoxal de l'autarcie dans l'œuvre de Rousseau et notamment dans ses fictions. Le rôle du philosophe, à l'instar du législateur, serait donc de substituer aux représentations collectives hétéronomes une nouvelle norme de rationalité. Cet apprentissage passe par une représentation autonome, voire autarcique de l'individu ou de la communauté, et renvoie, au-delà de l'une des figures traditionnelles de l'enseignement économique, Robinson Crusoé, au statut de la fiction dans la conceptualisation de la rationalité économique.

L'autorité morale de l'opinion ne permet pas aux hommes d'agir en fonction de leur véritable intérêt (*amour de soi*). En se soumettant à une autorité morale économique, sensée justement représenter les inégalités naturelles, les individus se déterminent en fonction d'un impératif de reconnaissance et de pouvoir que leur assure par ailleurs la richesse (*amour-propre*). L'état policé, décrit dans le premier puis second *Discours*, et dans lequel les individus agissent dans une apparente indépendance, ne doit sa stabilité qu'à l'hypocrisie et aux normes de civilité de ladite civilisation<sup>59</sup>. Le passage de l'état de nature à l'état civil induit, en revanche, un transfert de l'autorité morale de la richesse vers la loi. Celle-ci assure un statut identique à tous les citoyens. Les hommes, qui, dans l'état de nature, sont tous égaux devant la Nature, sont à l'état civil tous égaux devant la loi : « Chacun de nous met en commun sa personne et toute sa puissance sous la suprême

---

<sup>59</sup> J. Starobinski, *Jean-Jacques Rousseau: la transparence et l'obstacle ; suivi de Sept essais sur Rousseau*, Paris : Gallimard, 1989, p.165.

direction de la volonté générale ; et nous recevons en corps chaque membre comme partie indivisible du tout. »<sup>60</sup>

Cette égalité réinventée donne à chacun une identité commune qui transcende les identités personnelles forgées dans la différence<sup>61</sup>. Dépendants de la nature dans l'état de nature, les hommes restaurent leur liberté en s'extrayant des dépendances personnelles pour se soumettre collectivement et absolument à la loi<sup>62</sup>. C'est donc paradoxalement dans la réification et la sacralisation de la loi que dépend la liberté civile des hommes. Le législateur représente l'autorité capable d'établir la conformité des intérêts particuliers à l'intérêt de la nation, de faire « régner la vertu »<sup>63</sup>. La liberté civile, en abolissant le système de dépendance arbitraire auquel étaient soumis les hommes, leur permet donc de retrouver une forme d'autodétermination et de rationalité<sup>64</sup>.

Que devient l'échange dans l'état civil légitime ? Ou, plus exactement, quel rôle est-il amené à jouer dans une économie dans laquelle la loi représente la seule autorité morale légitime ? Rousseau ne s'est pratiquement pas exprimé sur le sujet, même si, on le sait, la

---

<sup>60</sup> J.J. Rousseau, *Œuvres complètes*, tome III : *Du Contrat social - Écrits politiques*, Paris : Gallimard, 1964, p.361.

<sup>61</sup> M. Viroli, *Jean-Jacques Rousseau and the 'Well-Ordered Society'*, Cambridge: Cambridge University Press, 1989, p.191.

<sup>62</sup> J.F. Spitz, *La liberté politique: essai de généalogie conceptuelle*, Paris : PUF, 1995, pp.368-69.

<sup>63</sup> J.J. Rousseau, *Œuvres complètes*, *op.cit.* p.252.

<sup>64</sup> Dans un contexte civil, la liberté ne peut être associée à la notion d'indépendance, ainsi que le précise Rousseau lui-même dans ses *Lettres écrites de la Montagne* (Iem, Viii, p. 841). La vraie liberté ne peut se résumer à une absence factuelle de contrainte, elle réside dans la garantie de l'absence de contrainte. La tradition républicaine ne s'oppose pas à la notion de liberté négative, elle compose avec. Philip Pettit et Quentin Skinner ont largement contribué ces dernières années à défendre le concept de liberté républicaine, entendu comme une non domination (par opposition à la non-coercition de la liberté négative). P. Pettit, *Republicanism: A Theory of Freedom and Government*, Oxford : Clarendon Press.

seconde partie des *Institutions politiques*, dont est issu le *Contrat social*, et qui n'a jamais été rédigée, devait traiter du commerce.

Dans un état légitime et vertueux, la médiation de l'échange n'est plus, en théorie, nécessaire à l'institution d'un lien social. L'état civil, en assurant par la loi, l'identité collective et la liberté des individus, permet donc théoriquement de restaurer la rationalité et la légitimité de l'échange. En devient-il pour autant libre ? Trois considérations nous amènent à penser que la rationalité économique ne se conjugue pas nécessairement pour Rousseau avec la liberté de l'échange. Tout d'abord, il est important de rappeler que, contrairement à l'amour-propre, l'amour de soi est une abstraction. Il représente une norme rationnelle de l'intérêt individuel, mais n'a pas de dimension réelle ou historique. Le contrat social s'appuie, comme la loi, sur une conception autonome de l'intérêt (*amour de soi*) mais il ne signifie pas que les individus agissent comme tel. Leurs motivations restent socialement déterminées. Rousseau ne rappelle-t-il pas, dès la première phrase du *Contrat social*, qu'il prend « les hommes tels qu'ils sont, et les loix telles qu'elles peuvent être »<sup>65</sup>? Les motivations de l'échange ne sont donc pas modifiées par la loi, seules ses pratiques peuvent l'être.

En conséquence, la loi est légitime lorsqu'elle encadre l'échange dans les limites que lui impose l'intérêt. Elle l'interdit ou le contraint lorsque les individus, emportés par l'envie et la vanité, cèdent à des désirs qui vont manifestement à l'encontre de l'intérêt commun et de leur intérêt particulier. C'est le cas notamment du commerce du luxe. La légitimité de

---

<sup>65</sup> J.J. Rousseau, *Œuvres complètes*, *op.cit.* p.351.

l'échange dépend donc de sa rationalité économique, elle-même assurée par la loi. La liberté de l'échange est garantie par la loi en même temps que l'autonomie morale des citoyens. Il y a donc, pour reprendre, sur ce point l'analyse de Jean Mathiot, une prééminence de la rationalité politique sur l'économique. L'architecture de la politique, selon son expression, désigne conjointement la capacité supérieure de la loi sur la richesse à instituer un ordre social, ainsi que sa supériorité heuristique. Non seulement la solution la plus juste est la plus appropriée, explique Jean Mathiot, mais elle peut également s'exprimer en terme d'intérêt<sup>66</sup>. L'existence de peuples historiquement constitués représente, ensuite, un véritable frein à l'institution d'une nouvelle norme de rationalité. Leurs lois et leurs coutumes, bien que souvent injustes et illégitimes, ont été acceptées et enseignées de générations en générations. Elles ont *de facto* participé à la formation d'une représentation collective de l'intérêt (*amour-propre*), lui-même légitimé par certains philosophes, Mandeville notamment. Vouloir réformer ces croyances est une entreprise, non seulement difficile, mais parfois aussi vaine et dangereuse<sup>67</sup>. Le peuple le plus propre à la législation est donc celui qui « n'a point encore porté le vrai joug des loix ; celui qui n'a ni coutumes ni superstitions bien enracinées »<sup>68</sup>. La Corse semble être l'un des seuls pays à répondre à l'ensemble des conditions posées par Rousseau<sup>69</sup>. Mais même, dans ce dernier cas, l'établissement d'une véritable citoyenneté est compromis par le poids des représentations morales de l'échange. Les préjugés des pays riches y sont en effet déjà

---

<sup>66</sup> J. Mathiot, « Politique et économie chez Jean-Jacques Rousseau », dans J. Boulad-Ayoub, I. Schulte-Jenckhoff et P-M. Vernes (dir.), *Rousseau. Anticipateur-retardataire*, Paris : L'Harmattan, 2000, p.28.

<sup>67</sup> J.J. Rousseau, *Œuvres complètes, op.cit.* p.385.

<sup>68</sup> J.J. Rousseau, *Œuvres complètes, op.cit.* p.390.

<sup>69</sup> Outre l'absence de lois, l'indépendance, l'état de paix, la taille et l'absence de richesses sont des éléments fondamentaux pour l'établissement d'un peuple nouveau, J.J. Rousseau, *Œuvres complètes, op.cit.* p.391.



répandus, « préjugés qu'il faut combattre et détruire pour former un bon établissement »<sup>70</sup>. Les hommes sont égaux de droit, mais restent inégaux dans l'opinion. Les représentations collectives de l'échange restent donc empruntées d'une symbolique passéiste. Des mesures de restriction, comme la limitation de la circulation de la monnaie, semblent donc d'abord répondre à des mesures de prudence et de pragmatisme politique. Elles participent également d'une volonté de réforme des représentations individuelles et collectives de l'intérêt.

Enfin, l'influence de la loi n'est, par définition, possible que dans le cadre de la nation. Les échanges internationaux y échappent nécessairement. Dotés d'une personnalité morale, les corps politiques sont, en effet, eux-mêmes soumis au joug des passions dont ils sont censés préserver leurs membres. Au niveau international, les nations sont donc encore dans un état de nature. L'échange constitue d'abord, comme la guerre, un mode d'affirmation de pouvoir et d'identité des nations<sup>71</sup>. Cela se traduit, à l'instar des individus, par une distorsion de l'utilité et du rapport d'échange. L'état hybride qui caractérise alors les sociétés est donc pour Rousseau le pire de tous, car « vivant à la fois dans l'ordre social et dans l'état de nature, nous sommes assujettis aux inconvénients de l'un et de l'autre, sans trouver la sûreté dans aucun des deux »<sup>72</sup>. Les relations entre nations s'inscrivent dans une démarche collective (et non plus individuelle), conflictuelle et identitaire. L'échange international répond alors aux exigences de visibilité et de discrimination de l'amour-

---

<sup>70</sup> J.J. Rousseau, *Œuvres complètes*, *op.cit.* p.902.

<sup>71</sup> Dans ses commentaires sur les écrits de l'abbé de Saint-Pierre, Rousseau envisage cependant l'hypothèse d'une fédération des nations européennes seule capable de concilier la notion communautaire de l'amour de soi et l'enjeu universaliste du droit naturel. Cf. Canto-Sperber [2005], p. 148-152.

<sup>72</sup> J.J. Rousseau, *Œuvres complètes*, *op.cit.* p.610.

propre. L'agriculture et les productions de première nécessité sont dévalorisés, et *in fine* délaissés, au profit d'activités plus lucratives telles que le commerce du luxe. Laisser l'échange libre entre les nations constitue, par conséquent, une menace pour la conservation du corps politique et de ses membres. Or, en l'absence d'une communauté européenne, voire mondiale, l'autonomie morale ne peut être réalisée que dans le cadre d'une communauté réduite et elle-même autonome. Une liberté totale de l'échange, pour répondre à l'exigence d'harmonisation des intérêts, impliquerait au niveau politique un pacte social universel.

Existe-t-il cependant une échappatoire à la contrainte politique que nous venons d'évoquer? La première possibilité, évoquée dans le *Discours sur l'économie politique*, consiste à envisager, à la suite de Diderot<sup>73</sup>, la possibilité d'une volonté générale globale. La volonté d'un peuple unique constitué en tant que personne morale s'imposerait alors également à tous comme norme universelle de justice<sup>74</sup>. La possibilité d'un corps politique au niveau mondial, la *société générale*, est cependant difficilement envisageable dans le cadre théorique de la nation développé par Rousseau. L'idée d'une société générale est d'ailleurs fermement rejetée dans le *Manuscrit de Genève*<sup>75</sup>. Une association globale des hommes n'étant ni possible ni souhaitable, une association de peuples, c'est-à-dire une *confédération*, permettrait selon lui d'associer la force du lien social qu'offre le patriotisme à l'universalité de la loi.

---

<sup>73</sup> J. Proust J., *Diderot et l'Encyclopédie*, Paris: Albin Michel, 1995, pp.388-90.

<sup>74</sup> J.J. Rousseau, *Œuvres complètes*, *op.cit.* p.245.

<sup>75</sup> J.J. Rousseau, *Œuvres complètes*, *op.cit.* p.284.

Le premier enjeu de l'économie politique rousseauiste consiste donc à dépasser, par une confédération européenne, un cadre national qui se réduit invariablement à un retour à l'état de nature des corps politiques eux-mêmes. Mais, à l'instar des individus, les corps politiques ne peuvent s'inscrire dans une démarche politique fédératrice lorsqu'ils rivalisent d'orgueil et de vanité. Si à l'intérieur de la nation, l'hypothèse d'une rationalité économique « dominée » peut paraître acceptable, en revanche elle ne peut être étendue au-delà du cadre de la nation. La notion de rationalité, qui trouve son fondement dans l'état de nature, a cependant d'abord une dimension universelle.

#### b. LA RECEPTION DE ROUSSEAU

Pourquoi l'approche fondamentalement politique de l'économie n'a-t-elle pas reçue le crédit qu'elle méritait ? L'article *Economie* de l'*Encyclopédie* constituait cependant une belle vitrine. J'ai évoqué ci-dessus, et dans les travaux qui y correspondent, deux raisons possibles. La première, développée dans ma thèse, est politique. De la reconnaissance du statut scientifique de la nouvelle économie politique, telle que Turgot et Morellet la définissait, dépendait la bonne mise en œuvre des réformes libérales voulues par le gouvernement et certains groupes de pression. Morellet, on l'a montré, était un mercenaire du libéralisme économique. Mais il n'était pas le seul. Les membres du groupe de Gournay étaient recrutés pour traduire les ouvrages économiques anglais ou espagnols, favorables à la liberté du commerce, puis encouragés à rédiger leurs propres traités. Face à l'avalanche de publications 'économiques' dans les années 1750, l'article de Rousseau est passé quasiment inaperçu. L'article a, ensuite, été désavoué par ses éditeurs, qui non contents de commander un second article 'Oeconomie Politique', ont utilisé le référent

‘économie politique’ pour désigner les seuls articles de Quesnay. Un troisième facteur, enfin, a pu nuire à la diffusion de la pensée de Rousseau jusqu’à l’aube de la Seconde Guerre Mondiale : les attaques personnelles incessantes dont il a fait l’objet et, plus spécifiquement, son anti-rationalisme et ses liens supposés avec le romantisme germanique.

C’est dans ce cadre que j’ai mené une étude sur la réception de l’œuvre de Rousseau avant-guerre. Bien que cette étude se limite à une période assez récente, celle de l’entre-deux guerres (les autres périodes étant déjà bien couvertes<sup>76</sup>), elle est symbolique des tentatives de décrédibilisation dont ses théories ont fait l’objet, et qui in fine ont permis au discours économique classique de s’établir.

L’article *Rousseau et la renaissance classique française (1898-1933)*<sup>77</sup> montre à quel point l’image de Rousseau au début du siècle dernier est éloignée de celle que l’on a de lui aujourd’hui. Rousseau est alors honni et son œuvre dévalorisée. On lui reproche d’être un apôtre de l’individualisme (alors que dans la période suivante, on lui reprochera ses penchants totalitaires), et de remettre en cause les valeurs classiques de la France. Au-delà de son intérêt historique, le but de cette étude est d’insister sur un phénomène que les historiens sous-estiment encore trop, à savoir que toute lecture, voire toute étude universitaire, nécessairement reflète le contexte idéologique et politique de son auteur, et que donc ce qui est valable pour l’auteur étudié, ici Rousseau, l’est tout autant pour le commentateur ou pour l’historien.

---

<sup>76</sup> Voir notamment C. Spector, *Au prisme de Rousseau : usages politiques contemporains*, Oxford, Voltaire Foundation, vol. VIII, 2011.

<sup>77</sup> C. Salvat, “Rousseau et la « Renaissance classique » française (1898-1933)”, *Astérior, Philosophie, Histoire des Idées, Pensée politique*, 12, 2014 [<http://asterion.revues.org/2545>]

C'est une crise morale, tout d'abord, pour tous les conservateurs, catholiques ou traditionalistes, qui assistent au déclin des valeurs religieuses et au triomphe du dandysme. Elle fédère, par réaction, de nombreuses personnalités intellectuelles ou politiques telles que Georges Bernanos, Maurice Barrès, Paul Bourget, Charles Péguy et naturellement Charles Maurras. Tous voient dans la société moderne une décadence morale causée par la glorification de l'individualisme au détriment des valeurs de la religion, de la famille et de la nation. En 1902, Pierre Lasserre déjà désigne le romantisme comme responsable de la décadence morale et politique de la France<sup>78</sup>. C'est une crise sociale, ensuite, dans laquelle l'affaire Dreyfus révèle au grand jour – au-delà de l'antisémitisme latent de la société française – la perte de l'autorité morale de l'État, et plus généralement celle de toute hiérarchie sociale. Elle se traduit par une radicalisation antidémocratique (Maurras, Sorel) ainsi que par la violence croissante des discours envers les minorités religieuses (notamment juives et protestantes), ainsi que contre les étrangers ou « métèques », en particuliers contre les Allemands accusés d'avoir corrompu la culture classique de la France avec le romantisme. Le mot d'ordre est celui de « Renaissance », d'un retour aux valeurs traditionnelles et méditerranéennes de la France. C'est une crise politique, enfin, qui se caractérise par une vague d'antiparlementarisme boulangiste, puis royaliste, associée à une montée en puissance du nationalisme. La Ligue de la patrie française et l'Action française sont fondées en 1898 sur des positions explicitement nationalistes et antidreyfusardes<sup>79</sup>. Les royalistes sont à la fin du XIXe siècle encore relativement peu nombreux, et sont essentiellement composés de catholiques conservateurs.

---

<sup>78</sup> P. Lasserre, *Le Romantisme français. Essai sur la révolution dans les sentiments et dans les idées au XIXe siècle*, Paris, Société du Mercure de France, p. 311-312.

<sup>79</sup> En France, la fin du XIXe siècle est marquée, comme chacun le sait, par deux grands événements, la défaite de 1870 et l'affaire Dreyfus. Ces deux événements sont généralement cités pour expliquer

La véritable campagne de dénigrement contre Rousseau débute en 1907-1908. Elle est menée par deux grandes figures de l'extrême droite, Pierre Lasserre et Jules Lemaître. Le premier, rédacteur du journal *L'Action française*, est un proche de Charles Maurras. Le second est, avec Maurice Barrès, membre dirigeant de la Ligue de patrie française, fondée en 1898 en réaction à la création de la Ligue des droits de l'homme. Les auteurs romantiques sont accusés de faire de leur personne et de leurs sentiments l'objet unique de leur intérêt. Bien qu'appartenant à deux mouvements politiques différents, Lasserre et Lemaître vont mener une campagne commune contre Rousseau et le romantisme. De calibres très différents, *Le romantisme français. Essai sur la révolution dans les sentiments et dans les idées au XIXe siècle* et *Jean-Jacques Rousseau*<sup>80</sup>, leurs ouvrages respectifs, tous deux publiés en 1907, vont exercer une très forte influence jusqu'à la seconde guerre mondiale.

À l'approche de la « Grande guerre », la question de l'identité culturelle et religieuse de Rousseau prend une nouvelle dimension. Tous ses critiques s'accordent à voir en lui un « étranger » ou un émissaire de l'étranger. Le romantisme est, rappelle-t-on à l'envi, un mouvement littéraire de tradition allemande. Le second grief fait à Rousseau est d'ordre moral. En rejetant les valeurs catholiques et traditionnelles de la France d'Ancien Régime, et en laissant les individus s'abandonner à leurs instincts naturels, Rousseau aurait encouragé le vice. Mais ce qui intéresse le plus les détracteurs de Rousseau à l'époque, ce

---

le nationalisme très affirmé d'une grande partie des Français, y compris des intellectuels et des artistes, avant la première guerre mondiale. Claude Digeon montre cependant que la situation est plus complexe qu'il n'y paraît, et que la réaction nationaliste et surtout antigermanique apparaît beaucoup plus tardivement. La comparaison des enquêtes sur les relations intellectuelles franco-allemandes, publiées par le *Mercure de France* respectivement en 1895 et 1902-1903 et étudiées par Claude Digeon, est édifiante. C. Digeon, *La crise allemande de la pensée française (1870-1914)*, Paris, Presses universitaires de France, 1959, p. 463 et suiv.

<sup>80</sup> J. Lemaître, *Jean-Jacques Rousseau*, New York : McClure Co., 1907.

sont ses pathologies psychiques<sup>81</sup>. La psychiatrie est alors, faut-il le rappeler, tout juste implantée en France, et le « cas Rousseau » est souvent donné en exemple. Les adversaires du romantisme peuvent donc aisément appuyer leur thèse sur des rapports et articles médicaux consacrés à Rousseau. Le troisième grief fait à Rousseau est de nature politique. La philosophie romantique est, selon ses détracteurs, anarchiste et révolutionnaire. Maurras, on le sait, a régulièrement associé Rousseau aux « désordres » de la Révolution française. Il déclare notamment que « Rousseau a été [...] la cause formelle de la Révolution ; il en a été l'âme et le génie, excitant les petits, stupéfiant et endormant les grands, donnant à l'attaque révolutionnaire des forces, à la défense traditionnelle de la faiblesse »<sup>82</sup>. En 1912, cependant, ce n'est pas la Révolution mais bien l'anarchie qui concentre l'essentiel des attaques contre Rousseau<sup>83</sup>. Les circonstances sont alors propices. Les auteurs exploitent l'émotion soulevée par les braquages et les meurtres perpétrés depuis décembre 1911 par Jules Bonnot, Octave Garnier et Raymond Callemin, principaux membres du groupe anarchiste dit de la « bande à Bonnot ». Lors d'un discours prononcé à la Chambre des députés le 11 juin 1912, Maurice Barrès s'en prend ainsi au gouvernement radical-socialiste d'Armand Fallières qu'il accuse de soutenir les milieux anarchiste

Ce premier article a été complété par un second portant sur l'immédiat après-guerre et traite de l'image de Rousseau comme étant l'une des influences idéologiques les plus

---

<sup>81</sup> R. Laforgue, « Étude sur Jean-Jacques Rousseau », *Revue française de psychanalyse*, vol. 1, n° 2, 1927 ; étude reprise et développée dans R. Laforgue, *Psychopathologie de l'échec*, Paris, Payot, 1939 (chap. 9)

<sup>82</sup> C. Maurras, *L'Action française*, 19 juin 1912, cité par R. Trousson, « Un ennemi de Rousseau : Charles Maurras », dans R. Trousson, *Défenseurs et adversaires de Jean-Jacques Rousseau. D'Isabelle de Charrière à Charles Maurras*, Paris, Honoré Champion, 1995, p. 318.

<sup>83</sup> Voir, par exemple, « Le métèque Rousseau » [proclamation de « l'Action française », citations de Proudhon, Lemaître, Maurras], *L'Action française*, 28 juin 1912 ; L. Daudet, « La chienlit de Rousseau, Le Fou et les Singes », *L'Action française*, 30 juin 1912 ; M. Pujo, « À bas le métèque », *L'Action française*, 1er juillet 1912. Sur les lectures « anarchiques » de Rousseau en 1912, voir T. L'Aminot, *Images de Jean-Jacques Rousseau de 1912 à 1978*, Oxford, Voltaire Foundation, 1992.

importantes du fascisme et du nazisme<sup>84</sup>. Berlin, Russel et Talmon, pour ne citer qu'eux, n'hésitent ainsi pas à associer Rousseau à Hitler et le rendre indirectement responsable de l'Holocauste. C'est, en large partie, en réaction aux idées de Rousseau que Berlin élabore quelques années plus tard sa célèbre distinction entre liberté positive et liberté négative. Il y aurait selon ces philosophes une logique totalitariste dans le système de Rousseau. Cette croyance a d'ailleurs prédominée jusqu'à la fin des années 60, et la réhabilitation de Rousseau dans la tradition libérale par John Rawls. Je montre dans cet article que si Rousseau a été lu de la sorte dès les années 30, c'est parce qu'un certain nombre de ses concepts, et en premier lieu, celui de la volonté générale, ont été l'objet d'une lecture 'étatiste' et 'collective' héritée, pour partie, des philosophes idéalistes anglais (Bosanquet, Green) et des républicains français (Durkheim, Bouglé). Pendant la guerre, celle-ci a été largement défendue par certains collaborateurs socialistes, tels que Déat, fiers de présenter Rousseau comme l'un des premiers totalitaristes de l'histoire. Rares sont ceux qui, à l'instar de Halévy, Hayek ou Cobban, voient dès les années 30 le danger d'une lecture hégélienne de Rousseau. La question n'est donc pas de savoir dans quelle mesure Rousseau a influencé la doctrine totalitariste, mais bien de savoir dans quelle mesure le contexte intellectuel et politique des années 20 et 30 a 'impregné' les interprétations des philosophes anciens (dont Rousseau mais pas exclusivement) de leur 'penchant totalitariste'.

---

<sup>84</sup> C. Salvat, From Literature to Politics How Rousseau Has Come to Symbolize Totalitarianism, *Journal of Interdisciplinary History of Ideas*, 6, (2017), 11, p. 4:1–4:23.



On pourrait naturellement poursuivre et étendre ces recherches à d'autres périodes. Céline Spector l'a d'ailleurs fait<sup>85</sup>, et ses travaux montrent que l'on pourrait en tirer des conclusions similaires. Ces deux études suffisent, je crois, à illustrer mon propos. Une troisième étude sur les lectures républicaines de Rousseau pourrait néanmoins utilement compléter cet ensemble. Bien que Philip Pettit et Quentin Skinner n'aient jamais accordé beaucoup d'importance aux écrits de Rousseau (de façon assez surprenante d'ailleurs), la grille de lecture qu'ils proposent a été ensuite largement utilisée pour réinterpréter son œuvre<sup>86</sup>. Cette grille de lecture, qui a ceci de particulier qu'elle s'appuie explicitement sur une approche historique fortement contextualisée, n'échappe néanmoins pas au biais souligné dans les études ci-dessus. Son interprétation du contexte historique du texte est nécessairement déterminée par les connaissances, mais également, les idéologies de l'époque moderne. De ce point de vue, elle n'est pas moins déterminée que ne l'était celle de Maurras, de Berlin ou de Talmon.

Ces études confirment la théorie dite de l'*œuvre ouverte*, chère à Umberto Eco<sup>87</sup>, qui affirme que le sens du texte lui est toujours *in fine* donné par le lecteur. Le texte *prend* un sens dans une situation historique, politique et intellectuelle particulière, et il est à la fois impossible et superflu de chercher à savoir si ce sens-là est celui que l'auteur voulait lui donner. L'intérêt d'une telle démarche ne s'arrête pas à ces strictes considérations méthodologiques puisqu'il est alors possible de montrer l'influence de ces lectures contextualisées sur la formation de théories modernes.

---

<sup>85</sup> C. Spector, *Au prisme de Rousseau : usages politiques contemporains*, Oxford, Voltaire Foundation, vol. VIII, 2011.

<sup>86</sup> Voir par exemple, J.F. Spitz, *La liberté politique: essai de généalogie conceptuelle*, Paris : Presses Universitaires de France, 1995.

<sup>87</sup> U. Eco, *Lector in fabula*, Paris: Grasset, 1985.

Les doutes exposés ci-dessus sur la neutralité des approches contextuelles, associés aux conséquences de l'historicisation croissante de l'histoire de la pensée économique (à savoir la perte de sa fonction critique et l'absence de dialogue avec les économistes), m'ont conduit à rediriger mes recherches vers la philosophie économique, qui connaît un fort engouement depuis quelques années.



## Deuxième partie

Doutes méthodologiques et réorientation

disciplinaire

## 1. De l'histoire intellectuelle et de ses limites

### a) HISTOIRE DE LA PENSEE ECONOMIQUE ET HISTOIRE INTELLECTUELLE

J'ai rédigé ma thèse au Centre Auguste et Léon Walras de Lyon, qui était alors la seule unité mixte de recherche du CNRS spécialisée en histoire de la pensée économique en France. Le Centre Walras, créé en 1984, avait déjà acquis une solide réputation nationale et internationale en publiant une édition critique des œuvres complètes de Léon Walras et de son père Auguste, en partie grâce au fond de lettres et de manuscrits dont il était dépositaire. Bien qu'ayant choisi de travailler sur le dix-huitième siècle, et sur le phénomène d'émergence du discours économique, mes travaux ont été fortement influencés par les positions méthodologiques et les pratiques de recherche, centrées sur le travail d'archive, défendues par ce centre. Le centre Walras était alors le plus ardent défenseur d'une histoire de la pensée contextuelle ou intellectuelle, par opposition à l'histoire de la pensée conceptuelle ou rétrospective, de tendance plus anglo-saxonne, et dont Mark Blaug était le plus important représentant<sup>88</sup>.

Les débats méthodologiques en histoire de la pensée économique dans les années 1990 occupaient une part importante des communications et des publications et pouvaient être très clivants. En histoire des idées (et notamment des idées politiques), ce débat méthodologique avait eu lieu dans les années 1970, notamment à la suite de la publication de l'article de Quentin Skinner, *Meaning and the Understanding in the History of Ideas*, en

---

<sup>88</sup> M. Blaug, *Economic Theory in Retrospect*, Cambridge : Cambridge University Press, 1962-1997.

1969<sup>89</sup>. Skinner s'opposait alors à deux types d'approches du texte. La première approche, de type contextuelle, appréciée des marxistes, consistait à expliquer le texte par le contexte socio-économique de sa production, au risque d'en faire un simple produit du contexte dont l'auteur ne serait d'un vecteur. La seconde approche interprétative à laquelle Skinner s'opposait était celle représentée alors par le philosophe américain Arthur Lovejoy, fondateur du *Journal of the History of Ideas*, et qui consistait à s'interdire toute considération extérieure au texte pour en expliquer la signification. Skinner proposa alors une troisième approche, qui consistait, à considérer le texte comme une participation à un débat intellectuel auquel l'auteur prenait part. Les textes n'existent pas indépendamment les uns des autres, et l'étude de la production intellectuelle de l'époque est nécessaire à l'étude de tout texte. Pour comprendre le sens ultime d'un texte, il faut, selon Skinner, d'abord comprendre l'intention première de l'auteur. L'intention (politique) de l'auteur est donc centrale dans l'analyse de Skinner. Cette théorie fit de nombreux émules, notamment à Cambridge où il a enseigné jusqu'en 2008, et où s'est constitué un groupe que l'on appelle aujourd'hui l'école de Cambridge en histoire intellectuelle, et regroupant notamment J. G. A. Pocock, Peter Laslett, John Dunn, James Tully, Istvan Hont, David Runciman ou Raymond Geuss.

L'influence de cette école en histoire de la pensée économique a été très importante en France dans les années 1980, et a contribué à renouveler ce champ disciplinaire longtemps dominé par l'histoire des 'doctrines' économiques de Charles Gide et Charles Rist<sup>90</sup>. Cette évolution fut, dans un premier temps, très bénéfique. Tout d'abord, elle lui a permis de

---

<sup>89</sup> Q. Skinner, « Meaning and the Understanding in the History of Ideas », *History and Theory*, Vol. 8, No. 1 (1969), pp. 3-53.

<sup>90</sup> C. Gide et C. Rist. *Histoire des doctrines économiques, depuis les Physiocrates jusqu'à nos jours*, Paris : Larose et Tenin, 1909. L'ouvrage connut 6 autres rééditions jusqu'en 1947.

renouveler la discipline face à une histoire conceptuelle saturée. Ensuite, et ainsi que j'ai tenté de l'illustrer dans mes propres travaux, l'histoire intellectuelle a donné un nouveau souffle critique à l'histoire de la pensée économique. Elle a été, tout d'abord, l'occasion de redécouvrir un corpus négligé par les économistes et propre à remettre en cause l'objectivité de leur discours.

Elle lui a permis, ensuite, de régénérer sa fonction critique. L'histoire de la pensée économique a, depuis qu'elle est enseignée, toujours eu un rôle critique même si, à certaines périodes, elle s'est montrée plus virulente qu'à d'autres. Les circonstances mêmes de son apparition et de son développement en France dans les années 1860-70, c'est-à-dire à l'apogée du positivisme, semblent attester sa forte dimension politique<sup>91</sup>. L'édition d'un numéro spécial d'*Economies et Sociétés* en 2014 pour les trente ans de la série Pensée économique (PE), numéro que j'ai coordonné, a très bien mis en lumière l'importance du contexte politique dans le développement de l'histoire de la pensée économique en France au début des années 1980<sup>92</sup>. Son évolution méthodologique illustre également les transformations de sa fonction critique.

L'école de Cambridge a, je l'ai dit, fait de l'histoire intellectuelle l'outil principal de l'étude de la pensée. Elle appelle donc à une historicisation, au moins partielle, de l'herméneutique. Selon Skinner, les textes et les idées se répondent les uns aux autres, et se nourrissent les uns des autres. Mais, notamment en histoire des idées politiques, ils

---

<sup>91</sup> C. Salvat, "French Political Economy and Positivism (or how history of economic thought became mainstream economics)", *Iberian Journal of the History of Economic Thought*, vol4., n°2, 2017. <https://revistas.ucm.es/index.php/IJHE/article/view/58093>

<sup>92</sup> C. Salvat (ed.) "Trente ans d'histoire de la pensée économique, 1984-2014 », numéro spécial, *Economies et Sociétés*, série PE, n°51, décembre 2014 ; C. Salvat, « Rétrospective des articles publiés depuis 1984 », *Economies et Sociétés*, numéro spécial 'Trente ans d'histoire de la pensée économique', 51, 2014, pp. 2015-2043.

répondent également à un contexte historique. L'histoire intellectuelle ne peut donc se limiter à l'étude d'un corpus, et doit nécessairement intégrer les faits. Elle requiert donc une certaine expertise historique. A l'université de Cambridge, les cours d'histoire des idées politiques sont enseignés au sein de la faculté d'histoire, dont Skinner occupa la prestigieuse chaire Regius de 1996 à 2008. Ils sont mutualisés avec la faculté de science politique. Les cours d'histoire de la pensée économique, en revanche, sont exclusivement enseignés dans les facultés d'économie, lorsqu'ils sont enseignés. L'historicisation de ce champ disciplinaire ou sous-disciplinaire a donc posé un certain nombre de difficultés pratiques liées à la formation et aux méthodes de travail de ses chercheurs. C'est dans ce contexte-là, que le Centre Auguste et Léon Walras a pu se distinguer en formant des doctorants aux pratiques du travail d'archive. L'histoire de la pensée économique s'est, ensuite, considérablement professionnalisée au contact des historiens.

#### b) LIMITES DE L'HISTOIRE INTELLECTUELLE

Mais l'historicisation croissante de l'herméneutique a aussi ses revers. Son principal travers consiste à perdre de vue l'importance du texte et des idées, et de ne s'intéresser qu'au contexte, voire aux anecdotes biographiques, de l'auteur. Cette tendance historiciste, encouragée par la ligne éditoriale de certaines revues, telles que *History of Political Economy*, a contribué à isoler davantage encore l'histoire de la pensée économique de la science économique. Les historiens de la pensée économiques se sont trouvés dans une situation particulièrement inconfortable : soit ils continuaient leurs recherches en tant qu'économistes, auquel cas leurs travaux faisaient (assez justement d'ailleurs) l'objet de la critique des historiens, soit ils adoptaient une perspective réellement historique, mais



perdaient l'attention des autres économistes, et risquaient, par conséquent, de perdre (au moins en partie) leur fonction critique à l'égard de la théorie économique. Ce fut pour cette dernière alternative qu'opta le Centre Walras. Et c'est à celle-ci que je me suis spontanément rallié lorsque je choisis m'intéresser à un économiste aussi peu connu et aussi peu théoriquement influent que l'abbé Morellet. La philosophie économique, ainsi que je le montrerais ci-dessous, a permis d'une certaine manière de réinvestir le discours critique de l'histoire de la pensée économique.

En quoi une historicisation excessive de l'herméneutique peut-elle nuire à la finalité première de l'étude de la pensée ? L'historicisation de l'étude des sciences sociales, qu'il s'agisse d'économie ou de philosophie, s'est traduite par un déplacement progressif, mais inéluctable, des objets de recherche du cœur de la discipline (des principaux textes/auteurs/théories) vers sa périphérie, puis dans un second temps de sa périphérie vers sa marge. L'histoire intellectuelle a, en effet, besoin de sources textuelles ou matérielles toujours renouvelées pour avancer. Les premières générations d'historiens intellectuels disposent d'un avantage si elle a la chance de pouvoir travailler sur des inédits fondamentaux pour comprendre la pensée de l'auteur étudié. Mais ceux-ci sont inévitablement en quantité limitée. Les générations suivantes doivent se contenter de sa correspondance et de quelques morceaux de textes inédits secondaires. L'environnement social, politique voire familial peut également être scruté. Mais les sources historiques restent limitées et les nouveaux entrants sont amenés à se concentrer sur des auteurs ou des sujets de moindre importance. C'est précisément la raison pour laquelle, étudiant, j'ai choisi de travailler sur un auteur aussi méconnu que l'abbé Morellet : la Bibliothèque Municipale de Lyon possédait un immense fonds de manuscrits inédits encore largement

inexploités. Dans ce cas-là, comme dans beaucoup d'autres, c'est l'existence de la source historique, plutôt que son intérêt théorique, qui justifie le projet de thèse.

Dans le cas de ma thèse, cependant, l'étude d'un auteur secondaire se justifiait également au regard de la finalité de ma recherche, à savoir la compréhension des conditions de production et de diffusion de la pensée économique libérale française du XVIIIe siècle. La fonction critique de l'histoire de la pensée économique a, dans le cas présent, été assurée.

Un deuxième revers possible d'une historicisation croissante de l'étude de la pensée consiste, pour le chercheur, à perdre de vue la finalité de l'enquête. Lorsque l'existence de la source justifie l'enquête, on peut être tenté de ne pas se poser de question sur la valeur de la source et donc sur la finalité ultime de l'enquête. Tout écrit, ou tout sujet de thèse, ne se justifie pas nécessairement par son caractère inédit ou original. Le travail d'archive, que j'estime par ailleurs beaucoup, a parfois tendance à nous le faire oublier. Lors de ma thèse sur Morellet, j'ai été immergé pendant plusieurs années dans un fonds de plus de deux mille manuscrits inédits. Son catalogage a nécessité, à lui seul, près de six mois de travail. J'avais la possibilité, et les invitations à le faire ne manquaient pas, de passer le reste de ma vie de chercheur à les commenter et à les publier. Mais dans quel but ? Ma thèse, aussi imparfaite qu'elle était, défendait une thèse, celle de l'organisation institutionnelle de la production et de la diffusion de textes économiques destinés à légitimer la liberté de commerce au XVIIIe siècle. La publication de la totalité des manuscrits n'y aurait rien apporté de plus. Nombreux étaient d'ailleurs les manuscrits qui ne s'y rapportaient pas. J'ai, cependant, pris soin, dans le second volume de ma thèse, de publier l'inventaire complet des manuscrits à l'intention de ceux qui pourraient y trouver des éléments pertinents pour leur propre recherche. Cet inventaire est aujourd'hui en ligne et accessible

à tous. Mon rôle s'arrêtait là. Poursuivre l'étude de ces milliers de manuscrits, et publier des dizaines d'articles sur le sujet, n'aurait en revanche rien apporté de mieux. La tentation de publier pour publier, simplement parce qu'on dispose de la matière, doit être contenue. C'est la raison pour laquelle, en dépit de l'avantage comparatif que me procurait ce fond d'inédits, et de ma connaissance unique de celui-ci, j'ai préféré réorienter mes thématiques.

Il existe, enfin, un troisième revers possible à une historicisation excessive de l'étude de la pensée, celui du constructivisme. Le constructivisme est très répandu dans les sciences sociales, notamment en France. Son succès doit beaucoup aux thèses de Michel Foucault notamment qui a très bien mis en lumière la dimension sociale et politique des mentalités et des savoirs. Le travail de thèse que j'ai réalisé sur les conditions de production et de diffusion du discours économique libéral français s'inscrit d'ailleurs dans cette perspective. Les discours en sciences sociales sont, naturellement le produit d'un contexte intellectuel, social et politique donné. Cela ne signifie, cependant, pas qu'ils ne sont qu'un instrument de pouvoir, ou qu'ils n'ont pas de valeur épistémique ou aléthique. Le fait que les conditions culturelles, sociales et politiques aient été favorables à l'émergence d'un discours scientifique n'est pas synonyme du fait que ces mêmes conditions aient déterminé le contenu de ce discours. Il est naturellement possible que cela ait parfois été le cas, mais la généralisation est abusive. Ainsi le fait d'avoir été payé par une institution pour produire une œuvre (littéraire, artistique ou scientifique) ne suffit pas à remettre en cause la valeur de cette œuvre. L'appréciation de la qualité d'une œuvre ou d'un discours doit rester indépendante de notre connaissance de ses conditions de productions.

Le principal danger du constructivisme, c'est le relativisme, c'est-à-dire l'idée que lorsque les discours émanent de structures socio-culturelles différentes, ils ne sont pas comparables. Le relativisme est dit scientifique lorsqu'il s'applique aux théories scientifiques, y compris aux sciences dites 'dures', ou moral lorsqu'il s'applique aux théories éthiques. Le relativisme scientifique est, malheureusement, de plus en plus répandu en France comme aux Etats-Unis. En France, on peut l'observer dans l'opposition d'une partie de la population à la médecine allopathique ou à la vaccination. Un tiers des français ont régulièrement recours à l'homéopathie, faisant de la France le leader de l'industrie homéopathique dans le monde. A part la France, seule l'Inde a massivement recours à ce type de médecine. Une étude américaine menée en 2018 portant sur 140.000 personnes de plus de 15 ans originaires de 144 pays différents a montré que, de tous ces pays, la France était le pays dans lequel ses habitants faisaient le moins confiance à la vaccination (1 français sur 3 remet en cause leur efficacité)<sup>93</sup>. Selon ces derniers, les médecines dites alternatives (l'acupuncture, la phytothérapie, la naturopathie, l'homéopathie, l'ostéopathie, la réflexothérapie, la sophrologie, l'aromathérapie, l'hypnothérapie etc.) ont une légitimité scientifique comparable à celle de la médecine dite conventionnelle. Ces derniers auraient, cependant, l'avantage de ne pas promouvoir les intérêts économiques des compagnies pharmaceutiques.

Aux Etats-Unis, le relativisme scientifique prend une dimension plus religieuse qu'institutionnelle. Son principal adversaire est la théorie évolutionniste darwinienne. Les Américains sont, dans leur grande majorité, favorables à l'enseignement du dessein

---

<sup>93</sup> [https://www.sciencesetavenir.fr/sante/la-france-premier-pays-anti-vaccins-selon-une-etude\\_134629](https://www.sciencesetavenir.fr/sante/la-france-premier-pays-anti-vaccins-selon-une-etude_134629)

intelligent, autrement dit du créationnisme, en complément de la théorie évolutionniste. Une minorité demande que seule le dessein intelligent soit enseigné dans les écoles publiques. Au-delà des croyances religieuses, l'argument des créationnistes est qu'une théorie (le dessein intelligent) en vaut bien une autre (l'évolution) et que l'éducation nationale n'a pas à exprimer de préférence pour l'une ou pour l'autre. Ce mouvement n'est pas propre aux Etats-Unis et s'installe également progressivement en Europe. En 2004, le gouvernement Berlusconi a déposé une proposition de décret visant à abolir l'enseignement de l'évolutionnisme dans les écoles et collèges. Grâce à la prompt réaction des scientifiques, le projet a été abandonné. En 2006, le parlementaire britannique Andrew McIntosh rédige un projet de résolution du Conseil de l'Europe intitulé *Les dangers du créationnisme dans l'éducation*, rapporté par le parlementaire socialiste français Guy Lengagne. La proposition est refusée en première lecture (64 voix contre 46) le 26 juin 2007 au nom de la liberté d'expression. Une seconde mouture, beaucoup plus conciliante avec les croyances religieuses, sera finalement adoptée en octobre de la même année. Cette remise en cause, non de la théorie évolutionniste en elle-même, mais de sa supériorité scientifique par rapport à d'autres modes d'explication du monde est caractéristique du relativisme scientifique. On le retrouve également dans la remise en cause du phénomène du réchauffement climatique aux Etats-Unis.

Le relativisme n'est pas le produit de l'historicisation des sciences sociales mais celles-ci ont néanmoins beaucoup contribué à son développement dans la seconde moitié du vingtième siècle, notamment en France. Michel Foucault, Claude Lévi-Strauss, Jean Baudrillard, Jean-François Lyotard, Bruno Latour y ont joué un rôle important, même si parfois leur discours a été un peu caricaturé et appliqué à des domaines auxquels ils ne

s'appliquaient pas forcément (je pense notamment à l'extension du structuralisme anthropologique de Strauss au domaine épistémique et moral). Leur autorité morale reste encore très importante, et de nouvelles générations poursuivent leur travail. En dépit de leur qualité intellectuelle, ces travaux reposent sur des bases fragiles et sur une fausse évidence, celle selon laquelle la pluralité historique des discours est la manifestation de l'absence d'objectivité scientifique. Ce n'est naturellement pas toujours le cas. Mais, même lorsque l'absence d'objectivité scientifique est manifeste, cela ne peut absolument pas constituer une preuve que l'extériorité de l'objet, elle, n'existe pas. Or c'est précisément ce que concluent relativistes et constructivistes. Dans le cas des sciences sociales, le jugement sur la nature scientifique du discours, autrement dit sa scientificité, ne se fonde pas sur une évaluation du degré d'objectivité atteint (comment pourrait-on le faire ?), mais sur la qualité du processus d'objectivation.

Il est fort heureusement possible d'adopter une démarche historique sans tomber dans le piège du constructivisme ou du post-modernisme, et j'espère ne pas y être tombé moi-même, mais l'expérience que j'en ai eu, et les discussions que j'ai pu avoir avec nombre de mes collègues, m'ont également appris qu'il était facile de passer d'une approche contextuelle à une approche constructiviste, ou qu'à l'inverse le constructivisme en sciences sociales menait droit à l'histoire contextuelle. Je dois la réorientation thématique, méthodologique et disciplinaire de mes recherches à une réflexion épistémologiques plus approfondie, mais également à une prise de conscience éthique et à certaines déconvenues méthodologiques.

### c) LES RAISONS D'UNE REORIENTATION

Pendant des années, j'ai été convaincu que la clé d'un texte, d'une œuvre ou d'une pensée était à chercher dans le contexte de production de l'œuvre, et que le travail d'interprétation du texte se doublait nécessairement d'un travail d'historien. Je me suis donc rapproché des historiens et de leurs méthodes jusqu'à intégrer la faculté d'histoire de l'université de Cambridge, qui constituait alors pour moi le temple sacré de l'histoire intellectuelle. J'y ai côtoyé Quentin Skinner, Itsvan Hont, John Dunn et Michael Sonenscher, ainsi que les très nombreux chercheurs et doctorants qui travaillaient dans leurs traces. Je dois, tout d'abord, dire que malgré ma prise de distance, je considère leurs travaux d'une qualité exceptionnelle et qu'ils m'ont beaucoup impressionné. Mais, en l'espace de cinq ans à leur côté, et dans ce cadre institutionnel très particulier, j'ai également réalisé que l'étude du contexte des conditions de production du texte, qu'elles soient intellectuelles ou socio-économiques, est elle-même influencée par le contexte intellectuel/socio-économique du chercheur. Or, les historiens sont très réticents à reconnaître le rôle que le contexte joue sur eux et leurs travaux, celle-ci pouvant être interprétée comme une remise en cause de l'objectivité de leur travail, et dans le cas de l'école de Cambridge, l'horizontalité de leur approche.

A la différence de celle d'un état du monde, l'étude d'un contexte historique est partielle et partiale. Sont contextuels les éléments de cet état du monde que l'historien considère comme pertinent pour expliquer l'œuvre. La sélection des éléments considérés comme pertinents, à défaut des autres, détermine de fait l'interprétation de ladite œuvre. Le fait que certains historiens de la pensée considèrent comme pertinents les seules conditions socio-économiques alors que d'autres ne considèrent, au contraire, comme pertinentes

que les seules les productions intellectuelles de l'époque illustre déjà l'option théorique choisie par le lecteur. Mais ce ne sont pas là les seules options possibles. D'autres historiens de la pensée ne s'intéressent, par exemple, qu'aux éléments biographiques voire psychologiques de l'auteur. Parfois ces considérations sont mêlées, mais dans tous les cas, les éléments contextuels considérés comme pertinents pour expliquer l'œuvre sont choisis par son interprète. A cela s'ajoute le fait que ces éléments sont souvent eux-mêmes recontextualisés, y compris à l'égard de l'œuvre qu'ils sont justement censés éclairer. La contextualisation est donc toujours, consciemment ou non, biaisée.

En réexaminant mon propre travail, je me suis rendu-compte du caractère parfois prédéterminé de mes analyses. Non seulement, mes recherches étaient orientées pour mettre en évidence une fracture politique entre les défenseurs et les opposants à la liberté du commerce, mais elles s'appuyaient sur des études qui adoptaient la même approche et la même représentation du contexte intellectuel et politique. Là où j'avais initialement vu une concordance des analyses, se révélait un simple mécanisme d'autorenforcement des opinions. Le travail que j'ai mené sur les différentes lectures de Rousseau, et notamment celle de Maurras, a achevé de me convaincre de l'importance du contexte du lecteur ou du chercheur. A partir de ce moment-là, j'ai décidé d'abandonner toute approche historicisante du texte, conscient néanmoins de l'impact inévitable de mon environnement sur mes propres lectures.

Parallèlement, j'ai progressivement découvert l'importance du réalisme scientifique et moral. Le réalisme désigne un ensemble de théories métaphysiques et méta-éthiques, qui ont pour point commun de défendre l'idée selon laquelle il existe des faits objectifs, y



compris des faits moraux, et que l'objet de la science (ou de la philosophie) est de les identifier et d'en comprendre les mécanismes. Le fait que les sciences, et notamment les sciences sociales, soient en capacité de remplir leur fonction relève d'une question épistémologique différente. En d'autres termes, le fait que certaines sciences (sociales) ne parviennent pas à découvrir ces faits, ou plus exactement qu'on ne parvienne pas à s'entendre sur le fait qu'elles y parviennent ou non, n'invalide en rien la position métaphysique défendue. Le réalisme scientifique, bien qu'il ait été remis en cause par les positivistes logiques, est plus intuitif que le réalisme moral, mais les deux sont, de mon point de vue, étroitement liés. Deux philosophes contemporains ont beaucoup contribué à faire évoluer mes vues en la matière, Derek Parfit et Pascal Engel. J'aurais l'occasion, dans la suite ce mémoire, de développer plus en détail la théorie du fondamentalisme des raisons de Parfit. Je voudrais ici dire brièvement quelques mots de celle de Pascal Engel, de l'importance que j'accorde à l'éthique intellectuelle et de son articulation avec le réalisme scientifique.

Dans son dernier ouvrage, *Les Vices du Savoir. Essai d'Éthique Intellectuelle* (2019), Pascal Engel pose les bases d'une nouvelle forme d'éthique, l'éthique intellectuelle, et condamne ce qu'il désigne comme certains vices du savoir. Pascal Engel refuse cependant de faire de l'éthique intellectuelle une branche de l'éthique normative. Il affirme ainsi que les « normes constitutives de la croyance sont propres au domaine épistémique »<sup>94</sup>. Autrement dit, la recherche du savoir, de la vérité, est une finalité morale en soi et non pas seulement une valeur instrumentale dans une démarche éthique. Ne pas reconnaître le

---

<sup>94</sup> P.Engel, *Les vices du savoir. Essai d'éthique intellectuelle*, Marseille : Agone, 2019, p.546.

phénomène du réchauffement climatique et la responsabilité de l'industrie humaine dans son accélération est une faute épistémique. C'est aussi une faute morale dans la mesure où nous sommes responsables de nos croyances et de leurs conséquences. Dans la mesure où il met gravement en danger les générations présentes et futures, le relativisme scientifique, tout comme le constructivisme et plus généralement le post-modernisme, relève – pour moi – de la faute morale.

Engel distingue une éthique intellectuelle d'une éthique normative dans la mesure où ne relèvent que de la première des raisons de type épistémique. Si, par éthique intellectuelle, on se réfère aux questions d'éthique qui sont soulevées *dans* l'exercice d'une activité intellectuelle – telle la triche aux examens, le plagiat, ou le travail bâclé de l'expert – alors non, l'éthique intellectuelle ne mérite pas d'être distinguée de l'éthique en général, voire de l'éthique professionnelle. Si, par contre, par éthique intellectuelle, on se réfère aux questions éthiques soulevées *par* l'exercice d'une activité intellectuelle, alors il est possible de défendre l'idée selon laquelle les spécificités de cet exercice (par rapport aux actions ordinaires des agents) en font un objet d'étude en soi. Les questions éthiques soulevées *dans* l'exercice d'une activité intellectuelle (recherche scientifique, journalisme, enseignement etc.) ont trait à l'évaluation des actions relevant de ces activités (la publication d'un article, la préparation d'un cours etc).

Le fait que ces actions relèvent de ces activités ne justifie cependant pas qu'elles soient soumises à des principes d'évaluation spécifiques. En général, leur valeur morale est déterminée par l'importance de leurs effets sur le bien commun. Les questions éthiques soulevées *par* l'exercice d'une activité (recherche scientifique, journalisme, enseignement

etc.) ont, en revanche, trait à l'évaluation de l'activité elle-même ainsi que de ses conditions de réalisation. Or, il est clair que la justification morale d'une activité intellectuelle ne serait se réduire à une simple évaluation des conséquences générées par sa pratique. L'activité intellectuelle répond, en effet, à d'autres fins que celles de produire de l'utilité. Elle se justifie également par l'importance morale de connaître la vérité, ou tout du moins, de tout mettre en place dans ce but. La valeur d'une action intellectuelle particulière – en tant qu'action intellectuelle – dépend donc de sa conformité aux raisons qui font de l'activité intellectuelle une activité morale. Il existe donc des normes propres à l'enquête intellectuelle qui ne seraient être confondues avec des normes morales.

## **2) De la philosophie économique**

Mon regard, et surtout ma pratique de l'histoire de la pensée économique (ou philosophique), a beaucoup évolué depuis mon entrée au CNRS. Je ne l'abandonne, cependant, pas totalement car elle me paraît être constitutive d'une approche complète d'une question théorique. Elle a vocation d'enrichir, plutôt que relativiser, les débats contemporains. C'est dans cet esprit que j'ai écrit l'ouvrage *L'Utilitarisme*, qui sera publié aux éditions de La Découverte à l'automne 2020.

Cet ouvrage se veut être une synthèse de la pensée utilitariste en même temps qu'une réflexion sur les principaux débats théoriques qu'elle a suscités et suscite encore aujourd'hui. L'histoire de la pensée y est naturellement présente mais reste limitée de manière à pouvoir présenter et discuter les développements récents de cette théorie, notamment toutes les questions liées à la mesure et à la maximisation de l'utilité, son

articulation avec les théories du bien-être, son égalitarisme contesté ou encore tous les enjeux théoriques liés à sa problématique universalisation. La publication d'un tel ouvrage en français me paraît particulièrement importante car la théorie utilitariste est très mal connue en France et souffre d'une mauvaise image, souvent associée à une profonde mécompréhension de ce mouvement. De mon point de vue, c'est aussi une occasion intéressante de creuser la nature des relations entre philosophie et économie.

L'articulation entre l'économie et la philosophie est au centre de mes intérêts depuis de longues années. Elle a néanmoins beaucoup évolué ces dernières années, devenant progressivement moins historique et plus analytique, et davantage orientée vers les questions d'éthique. Cette évolution s'est faite en plusieurs étapes, mais mes travaux sur le paternalisme comportemental de Sunstein et Thaler y ont certainement beaucoup contribué.

#### A) CE QU'EST LA PHILOSOPHIE ECONOMIQUE

La philosophie économique désigne le recoupement existant entre l'économie et la philosophie. Celui-ci porte essentiellement sur les questions relatives à la rationalité ou à la moralité de l'action. Depuis vingt ou trente ans, on observe un intérêt croissant de la science économique pour les questions traditionnellement traitées par la philosophie, et notamment les questions éthiques. Les relations croisées entre discipline, et particulièrement entre économie et philosophie, sont très bénéfiques car elles permettent d'intégrer des questionnements et des concepts nouveaux dans chaque discipline. Parmi les économistes qui ont le plus contribué à ouvrir la science économique aux questionnements philosophiques, on peut notamment mentionner Amartya Sen.

La philosophie, au même titre que l'histoire, donne aux économistes la distanciation nécessaire pour développer une réflexion critique sur leur propre discipline. Son objet, cependant, n'est pas nécessairement le même. De par sa nature, l'histoire s'interdit généralement l'étude des théories ou des auteurs contemporains. Elle perd donc, au bénéfice de la rigueur sans doute, la possibilité d'engager directement avec les économistes. Ce n'a pas toujours été le cas, mais l'évolution très rapide de la discipline rend, ainsi que je l'ai indiqué plus haut, les comparaisons historiques plus difficiles. La philosophie économique permet ce dialogue avec le présent tout en n'excluant pas le passé. Sa nature fortement théorique la rend, cependant, peu propice aux approches contextuelles.

Institutionnellement parlant, la philosophie économique est un domaine encore nouveau en France. La *Revue de philosophie économique* fut créée par Alain Leroux en 2002 à Aix-en-Provence, avec lequel j'ai travaillé dix ans<sup>95</sup>. L'expression, quant à elle, est ancienne ; on la trouve déjà chez Baudeau<sup>96</sup>, et elle resurgit tout au long du 20<sup>e</sup> siècle. Daniel Villey fut, sans doute, le premier non seulement à avoir donné un véritable cours de philosophie économique mais en à avoir défendu le principe<sup>97</sup>. Villey conçoit la philosophie économique comme l'application de la philosophie au domaine de l'économie. Puisque la philosophie se divise, selon lui, en trois axes, l'ontologie, la gnoséologie et l'axiologie, la philosophie économique doit naturellement s'organiser autour de l'ontologie, de la

---

<sup>95</sup> A mon entrée au CNRS, en octobre 2002, je fus nommé au département de philosophie économique du GREQAM (Groupement de Recherche en Économie Quantitative d'Aix-Marseille, UMR 7316) à Marseille. J'y suis resté jusqu'en 2012.

<sup>96</sup> N. Baudeau, *Première introduction à la philosophie économique: ou analyse des états policés*, Paris : Didot, 1771.

<sup>97</sup> D. Villey, « Prolégomènes à l'enseignement de la philosophie économique », *Revue d'Economie Politique*, Mai-Juin 1959, pp. 362-414.

gnoséologie et de l'axiologie économiques. Par ontologie économique, Villey entend 'la signification *philosophique* de l'objet de l'économie politique'. Il définit la gnoséologie économique comme étant « l'examen des fondements et des postulats philosophiques de la science économique » et appelle de son vœu une axiologie économique « pour définir et pour apprécier les finalités autonomes que promeut la vie économique et que dégage la science économique»<sup>98</sup>.

Cette approche est rejetée par Leroux et Marciano pour qui la philosophie économique « ne désigne pas davantage une philosophie appliquée à l'économie qu'une économie à prétention philosophique. Elle est constituée par les travaux d'économistes et de philosophes qui, pratiquant leur propre discipline, trouvent des problématiques communes. Ainsi abordé, le contenu de la philosophie économique n'est donc plus à inventer mais seulement à découvrir»<sup>99</sup>. L'approche de Leroux et de Marciano est quelque peu problématique dans la mesure où elle refuse à la réflexion philosophique toute fonction critique<sup>100</sup>. Or, il me semble, au contraire, que c'est là sa principale force, et que c'est à cette fonction critique, en partie perdue par l'histoire de la pensée économique, que l'on doit son important développement ces quinze dernières années.

---

<sup>98</sup> D. Villey, « Prolégomènes à l'enseignement de la philosophie économique », *op. cit.*, p.401.

<sup>99</sup> A. Leroux, A. Marciano, *La philosophie économique*, Paris: Presses Universitaires de France, collection Que sais-je?, 1998, pp.5-6.

<sup>100</sup> Les auteurs sont, en effet, plus enclins à mettre en doute la rigueur conceptuelle des philosophes que celles des économistes : «A bien des égards, l'approche de la science prend le contre-pied de celle de la philosophie. Alors que la première trouve généralement l'assise de son discours dans la plus profonde intériorité, la science ne veut faire confiance qu'à la plus parfaite extériorité, celle qu'elle trouve dans les faits. La philosophie est personnelle, la science impersonnelle. » A. Leroux, A. Marciano, *La philosophie économique*, Paris: Presses Universitaires de France, collection Que sais-je?, 1998, p.36.

Dans de nombreux cas, cependant, la philosophie économique échoue à véritablement nouer le dialogue avec la philosophie et se contente d'investir un domaine, celui de l'éthique par exemple, sans se doter du matériel conceptuel qui lui correspond. C'est, je crois, la grande faiblesse de la théorie de Sunstein et Thaler, qui est aujourd'hui l'un des thèmes les plus discutés dans les colloques de philosophie économique. J'ai également récemment émis de fortes réserves sur une étude intitulée « The Moral Machine Experiment », parue dans le journal *Nature*. Cette étude me paraît, en effet, assez caractéristique des déviations que l'on peut observer aujourd'hui. Elle rend compte d'une expérience en ligne menée sur grande échelle et proposant différents scénarios virtuels impliquant une voiture autonome, ses passagers et des piétons. L'enjeu est de tester les intuitions morales des internautes et d'identifier dans chaque scénario quels individus sacrifier en fonction de leur âge, sexe, statut social, état de santé etc. et de construire, à partir de ces données, des règles morales 'sûres' que l'on puisse légitimement imposer. En l'absence de toute réflexion méta-éthique sur l'existence même de telles normes, et sur la nature externe ou interne de leur justification, les auteurs sont incapables de véritablement tirer bénéfice de leurs données. Plus grave encore, en se proposant de donner des fondements scientifiques à l'éthique, ils dénie toute expertise éthique aux philosophes et contribuent à discréditer, non seulement la philosophie, mais également le reste des sciences humaines et sociales. Pour être valable et bénéfique à tous, l'interdisciplinarité doit d'abord reconnaître l'expertise des autres. C'est précisément cette conviction qui m'a conduit à rejoindre un laboratoire de philosophie pour poursuivre mes travaux.

En poste depuis mai 2013 au laboratoire Triangle, associé à l'Ecole Normale Supérieure de Lyon, j'ai demandé et obtenu ma mutation au centre Gilles Gaston Granger au 1<sup>er</sup> mars

2018. Cette mobilité géographique correspond avant tout à la mobilité thématique et disciplinaire exposée plus haut. Bien que le laboratoire Triangle soit interdisciplinaire, les possibilités de collaboration avec les philosophes étaient, de fait, très limitées. Entre 2013 et 2018, je suis néanmoins parvenu à mettre en place un axe 'Philosophie économique' au sein de l'équipe des économistes, essentiellement composée d'historiens de la pensée économique. Je me suis également beaucoup investi dans l'activité du laboratoire, étant membre nommé du conseil de laboratoire entre 2015 et 2018. J'ai enfin organisé un certain nombre d'évènements scientifiques, dont le 4<sup>e</sup> Colloque International de Philosophie Economique (2018), et le séminaire annuel de philosophie économique. Peu de temps après mon arrivée à Triangle, s'est posée la question de la place de la philosophie économique au sein du centre. Personnellement j'étais favorable à ce que celle-ci puisse se développer au sein de l'équipe des philosophes, ou tout du moins, en étroite relation avec celle-ci. Un choix différent a été fait, et il a été décidé qu'un sous-axe 'Philosophie économique' serait créé au sein de l'axe Economie. Ce choix n'est pas dénué de sens puisque, au moins en France, la philosophie économique s'est très largement développée dans les structures institutionnelles de l'histoire de la pensée économique que ce soit l'Association Charles Gide pour l'Etude de la Pensée Economique, les revues ou les séminaires d'histoire de la pensée économique. Il existe donc une très forte perméabilité entre ces deux sous-disciplines. Cela a certains avantages (cela renouvelle l'HPE en même temps que cela donne une profondeur historique à la PE), mais aussi certains inconvénients.

Le premier d'entre eux est de diminuer les possibilités de dialogues et de coopération avec les philosophes. Les historiens de la pensée économique bénéficient, à l'instar des philosophes, d'un réseau institutionnel très bien établi, et chacun communique et publie



dans ses propres institutions. Une brève comparaison des travaux publiés de part et d'autre sur des sujets similaires montre la faible interpénétration des différents corpus scientifiques. Je considère donc que pour faire de l'interdisciplinarité une réalité, celle-ci doit impliquer un transfert institutionnel. C'est la raison pour laquelle, lorsque je travaillais en histoire de la pensée économique, j'ai été travaillé dans le département d'histoire de l'Université de Cambridge (2006-2011), et c'est la raison pour laquelle j'ai aujourd'hui demandé à intégrer une équipe de philosophe à Aix-en-Provence. Fusionner l'histoire de la pensée économique et la philosophie économique présente, de mon point de vue, un second inconvénient. Traditionnellement, l'histoire de la pensée économique en France aime à se présenter comme une approche hétérodoxe de l'économie. Elle considère le discours économique comme étant l'expression d'un rapport de pouvoir, dans le cas présent l'expression de la domination capitaliste. Tous les points de vue sont naturellement acceptables et je préfère personnellement éviter de rentrer dans tout débat de nature idéologique. En revanche, je ne partage pas l'idée selon laquelle il n'y a pas de connaissance objective, y compris en éthique. Ce discours conduit nécessairement au relativisme scientifique, à l'idée selon laquelle il n'y a pas de propositions vraies ou fausses, mais simplement des opinions également respectables. Cette attitude peut s'avérer très dangereuse. J'ai déjà indiqué que, bien que très attaché à l'histoire des idées, j'ai préféré m'éloigner de toute histoire de la pensée qui pouvait justifier cette attitude. Je ne souhaite donc pas voir aujourd'hui la philosophie économique emprunter le chemin emprunté hier par l'histoire de la pensée économique. Pour toutes ces raisons, il m'a semblé préférable d'intégrer une équipe de philosophes pour poursuivre mes travaux.

Le choix du centre Gilles Gaston Granger, attaché à l'université Aix-Marseille, a été motivé par plusieurs considérations. La première d'entre elles, et ainsi que son nom l'indique, est

son interdisciplinarité et sa spécialisation en histoire des sciences. Le centre, nommé CEPERC jusqu'au 1<sup>er</sup> janvier 2018, a une longue tradition en épistémologie et philosophie des sciences, notamment en mathématique et physique, mais plus récemment aussi en biologie et sciences de la terre. La science économique y a également historiquement tenu une place importante, grâce aux travaux de son fondateur, tout d'abord, qui a aujourd'hui donné son nom au centre, puis grâce à ceux de Pierre Livet et de son équipe. Depuis quelques années, et sous l'impulsion de son nouveau directeur, Pascal Taranto, le centre a été profondément réorganisé. Il est aujourd'hui composé de deux axes, l'axe 1 en 'Histoire et philosophie des sciences', et l'axe 2 'Histoire de la philosophie et ontologies du présent', que je coordonne avec Michel Le Du. L'objet de ce second axe est, entre autres, de développer les analyses de nature philosophique ayant trait à la société et à son évolution. C'est dans cet axe que s'inscrivent mes recherches en philosophie économique, en éthique et en éthique appliquée. Au-delà de la poursuite de mes travaux personnels, ma présence dans ce centre me permet de mettre en place des collaborations fructueuses, avec Michel Le Du par exemple avec lequel j'organise en 2021 un colloque sur le rôle des émotions.

Pour terminer, je propose maintenant présenter deux thèmes de recherche sur lesquels j'ai eu l'occasion de travailler en philosophie économique, la justification du paternalisme et l'importance de la notion d'identité personnelle pour la rationalité. Ils représentent deux démarches alternatives de la philosophie économique. Dans le cas du paternalisme, la rencontre entre la philosophie et l'économie est initiée par les économistes qui voient dans le paternalisme une implication logique de leurs travaux sur la rationalité. Dans le cas de l'identité personnelle, au contraire, ce sont les philosophes, dont la conception de l'identité remet potentiellement en cause les standards économiques de rationalité, qui engagent le dialogue avec les économistes.

## B) LE PATERNALISME COMPORTEMENTAL

En 2003 paraît un article qui va profondément influencer mon travail de recherche, et dont l'importance théorique (mais également politique) ne se révélera que quelques années plus tard. Il s'agit de *Libertarian Paternalism Is Not an Oxymoron*<sup>101</sup> de l'économiste Richard Thaler et du juriste Cass Sunstein.

Pour bien comprendre les fondements du paternalisme comportemental, il est important de revenir brièvement sur les principales conclusions de l'économie comportementale. Les résultats des travaux des économistes comportementaux, auxquels appartient Richard Thaler, peuvent schématiquement être regroupés en deux catégories : 1) les études empiriques ou expérimentales prouvent que les actions des agents économiques ne satisfont pas aux normes de rationalité utilisées dans la théorie du choix rationnel 2) ces mêmes études tendent à montrer que certains biais comportementaux ou contextuels (tels que le *framing* ou l'*endowment effect*) contribuent à expliquer de la sous-optimalité des agents économiques. A la différence des théories standard du paternalisme, qui trouvent leurs justifications dans l'irrationalité des préférences individuelles (irrationalité de préférer ne pas porter un casque en conduisant une moto ou de ne pas épargner pour sa retraite), le paternalisme comportemental préfère voir dans l'irrationalité d'une action un effet de contexte qui n'engage pas nécessairement les facultés cognitives de l'agent. Il s'agit là d'un point fondamental puisqu'il nous autorise à penser que, placé dans une situation moins exposée aux biais comportementaux, le même individu agirait de manière plus conforme à son intérêt. L'hypothèse est certes spéculative, mais elle a un corollaire

---

<sup>101</sup> C. Sunstein, R. Thaler, "Libertarian Paternalism Is Not an Oxymoron", *The University of Chicago Law Review*, 70:4, 2003, pp. 1159-1202.

important : à condition de ne pas être exposé aux mêmes biais, n'importe quel autre individu peut être capable d'identifier les effets pervers du contexte sur la rationalité de l'agent. Prenons, tout d'abord, l'exemple de l'*endowment effect* que Richard Thaler a personnellement contribué à identifier<sup>102</sup>. Supposons deux individus, A et B, à qui l'on propose de choisir entre deux biens, une tasse et un stylo, de même valeur. L'expérience menée par Kahneman, Knetsch & Thaler (1990) montre que l'agent A, à qui l'on a réellement donné la tasse avant de débiter l'expérience, est deux fois moins susceptible de l'échanger que l'agent B, qui raisonne simplement in abstracto<sup>103</sup>. Modifions maintenant les termes de l'expérience, et supposons que le stylo proposé en échange de la tasse a une valeur supérieure à celle-ci. A et B devraient donc logiquement préférer le stylo à la tasse. Dans l'hypothèse où A continue cependant à préférer la tasse au stylo, on peut légitimement conclure que l'*endowment effect* est responsable de la sous-optimalité de l'action de A. Ce défaut de rationalité est facilement identifiable par ceux qui ont la chance de ne pas être soumis aux mêmes conditions de choix que l'agent A. Placé dans un contexte différent, l'agent A lui-même agirait probablement de façon différente.

L'implication théorique de cette conclusion sur le paternalisme est immédiate. Nul besoin donc que le paternaliste soit doté de facultés cognitives supérieures pour justifier de sa fonction. Sa position seule lui offre un avantage relatif. Du point de vue de la rationalité, aucun agent n'a donc supériorité cognitive absolue sur les autres. Ce point est absolument essentiel car il justifie l'abandon de l'Etat comme acteur privilégié du paternalisme au profit

---

<sup>102</sup> R. Thaler, "Toward a Positive Theory of Consumer Choice", *Journal of Economic Behavior and Organization*, 1(1), 1980, pp. 39-60.

<sup>103</sup> D. Kahneman, J. Knetsch and R. Thaler, "Experimental Tests of the Endowment Effect and the Coase Theorem", *Journal of Political Economy*, vol. 98 (6), 1990, pp.1325-48.

de l'individu. J'aimerais insister ici sur le caractère relatif de l'avantage de l'agent paternaliste. Nous sommes tous, plus ou moins, soumis à un certain nombre de biais comportementaux dont nous ne sommes généralement pas conscients. S'il nous est aisé d'identifier les biais comportementaux des autres, ainsi que leurs effets sur leur rationalité, il nous est en revanche quasiment impossible de nous soustraire aux nôtres par notre seule volonté. Le paternalisme comportemental peut donc être considéré comme un mécanisme de coopération spontané dans lequel les agents paternalistes partagent leur avantage positionnel avec ceux qui n'en bénéficient pas, et vice-versa.

L'importance qu'a prise la théorie dite des 'nudges' auprès de l'opinion publique et des instances politiques atteste du fait que les auteurs ont touché un point névralgique, et que l'on assiste en ce moment à un retournement idéologique, retournement auquel j'ai consacré un colloque interdisciplinaire à Lyon en 2015<sup>104</sup>. La thématique du paternalisme m'intéressait donc du point de vue historique mais également du point de vue théorique. La théorie de Sunstein et Thaler repose sur l'articulation logique de trois affirmations : a) les actions individuelles sont fortement influencées (voire déterminées) par un certain nombre de biais comportementaux ; b) ces biais, dont l'effet est généralement défavorable aux agents, peuvent être paramétrés pour, au contraire, renforcer la rationalité des actions individuelles ; c) il est souhaitable que les individus (appelés 'planificateurs') qui ont la possibilité de modifier les biais comportementaux le fassent et agissent dans l'intérêt de tous. Ces différentes propriétés, ainsi que l'ont montré Philippe Mongin et Mikäel Cozic,

---

<sup>104</sup> *Paternalism Redeemed : Old Ideals, New Realities*, 30-31 March, ENS de Lyon. Ce colloque regroupait des historiens (C. Armenteros, G. Claeys, O. Siméon), des économistes (C. Hédoïn, C. Salvat) et des philosophes (D. Scoccia, L. Bovens, J. Wilson). Il a été suivi par un second colloque, *Paternalism, Liberalism, Socialism : France and England, 19th C.*, organisé à Oxford les 3 et 4 juin 2016 en collaboration avec L. Frobert et M. Drolet.

sont généralement confondues dans la notion de *nudge*<sup>105</sup>. Il est important qu'elles soient distinguées, autant par rigueur définitionnelle, que par égard aux différentes implications normatives qu'elles peuvent avoir.

La première et la deuxième allégation relèvent de la simple théorie économique. L'existence de biais comportementaux et leurs mécanismes sont maintenant bien connus grâce aux nombreux travaux d'économie comportementale, que Thaler a lui-même contribué à développer<sup>106</sup>. C'est au niveau de la dernière allégation que peut intervenir le philosophe. Elle contient deux éléments de débat, la nature et la désirabilité de l'action correctrice. Il existe trois réponses possibles aux biais comportementaux. La première est l'action paternaliste, individuelle et volontaire. C'est celle défendue par les auteurs. La deuxième est l'action légale, coercitive et démocratique. Considérée, peut-être à tort, de paternaliste<sup>107</sup>, elle représente l'alternative traditionnelle que Sunstein et Thaler essaient de dépasser en lui opposant une forme de paternalisme dite libertarienne. L'essentiel des débats s'est concentré sur la seule question de la désirabilité de l'action paternaliste, indépendamment de celle de ses alternatives.

Une action peut être considérée comme désirable lorsqu'elle est rationnelle et/ou morale. Les auteurs considèrent que la rationalité de l'action qu'ils proposent, à savoir l'action paternaliste, est démontrée. En réalité, elle ne l'est pas ou, du moins, pas totalement.

---

<sup>105</sup> P. Mongin et M. Cozic, "Rethinking nudge: not one but three concepts", *Behavioural Public Policy*, 2(1), 2018, pp. 107-124.

<sup>106</sup> G. Loewenstein, D. Prelec, "Anomalies in Intertemporal Choice: Evidence and an Interpretation", *The Quarterly Journal of Economics*, 107:2, 1992, pp. 573-597; A. Tversky, D. Kahneman, "The Framing of Decisions and the Psychology of Choice", *Science*, 211: 4481, 1981, pp. 453-458; A. Tversky, P. Slovic, D. Kahneman, "The Causes of Preference Reversal", *The American Economic Review*, 80:1, 1990, pp. 204-17.

<sup>107</sup> G. Dworkin, "Paternalism", *The Monist*, 1, 1972, pp. 64-84; J. Feinberg, "Legal Paternalism", *Canadian Journal of Philosophy*, 1:1, 1971, pp. 106-24.

L'essentiel de leur argument est destiné à défendre le caractère moral de l'action paternaliste. Mais même de ce point de vue, celui-ci n'est pas totalement convainquant.

Le paternalisme peut être soumis à trois objections morales : son caractère liberticide, sa nature anti-égalitariste et sa légitimité. La première critique est la plus commune, et c'est à celle-ci que Sunstein et Thaler se proposent de répondre sans parvenir toutefois à prendre toute la dimension du problème. Je reviens sur ce point un peu plus bas. Les deux dernières objections ne sont pas abordées par les auteurs. Elles sont néanmoins fondamentales. Considérons, tout d'abord, la dernière. La question de la légitimité de l'action ne se pose pas dans le cas du *self-control* : l'agent a le droit moral de se contraindre lui-même. Elle ne se pose plus véritablement non plus dans le cas d'une constitution démocratique, où les citoyens se sont volontairement soumis à la règle de la majorité. En revanche, elle se pose dans le cas de l'action paternaliste, dans laquelle une personne prend l'initiative de contraindre les choix d'une autre personne. Le caractère bienveillant de cette initiative l'exempte-t-elle du principe de respect de l'autonomie individuelle ? Les auteurs semblent penser que c'est le cas mais ne justifient pas leur point de vue. La seconde objection morale au paternalisme a été développée notamment par Elisabeth Anderson, qui voit dans cette pratique une atteinte au principe de l'égalité de respect que l'on doit aux personnes<sup>108</sup>. Se permettre d'intervenir pour le bien d'une tierce personne, c'est considérer – selon elle – qu'elle n'est pas à-même de décider pour elle-même, ou qu'elle l'est dans une moindre mesure que vous. L'argument est difficilement contestable. Néanmoins, il est plus adapté aux formes traditionnelles de paternalisme qu'au

---

<sup>108</sup> E. Anderson, "What Is the Point of Equality?", *Ethics*, 109:2, 1999, pp. 287-337.

paternalisme comportemental, ainsi que j'ai essayé de le montrer<sup>109</sup>. Une approche plus approfondie des aspects philosophiques de leur théorie aurait permis aux auteurs de répondre à cette objection. La seule objection dont ils ont tenu compte, ainsi que je l'ai dit, c'est celle relative au caractère coercitif du paternalisme.

Y ont-ils répondu de façon appropriée ? L'argument qu'ils défendent est que le paternalisme comportemental, autrement dit celui qui a pour objet de débiaiser le comportement des individus, n'est pas réellement contraignant dans la mesure où les agents ont toujours la possibilité, s'ils le souhaitent vraiment, de ne pas s'y conformer. Il s'agit simplement, pour reprendre leur expression, d'un 'coup de pouce'<sup>110</sup>. De fait, et bien qu'ils n'approfondissent pas cette question, ils introduisent l'idée de gradation dans la volonté de l'agent, aussi appelée *voluntariness*<sup>111</sup>. Cette notion permet de distinguer, au moins sur le plan théorique, les préférences individuelles autonomes, dont on peut légitimement tenir compte, des préférences hétéronomes que l'on peut légitimement ignorer. Ce concept d'inspiration kantienne a notamment été utilisé pour défendre une lecture pro-paternaliste de John Stuart Mill<sup>112</sup>. Il est cependant incompatible avec le principe de base de l'économie comportementale, principe selon lequel nos préférences sont le produit de notre environnement<sup>113</sup>. L'introduction de cette notion réduit, par ailleurs, à néant la distinction entre un paternalisme légal coercitif et un paternalisme

---

<sup>109</sup> C. Salvat, "Behavioral Paternalism", *Revue de Philosophie Economique*, vol.15, n°2, 2014, pp.109-30.

<sup>110</sup> R.H. Thaler, C.R Sunstein, *Nudge. Improving Decisions About Health, Wealth, and Happiness*, New Haven & London: Yale University Press, 2008.

<sup>111</sup> Cette notion est centrale dans les travaux de Feinberg notamment, J. Feinberg, *Harm to Self: The Moral Limits of the Criminal Law*, New York: Oxford University Press, 1986.

<sup>112</sup> J'ai néanmoins tenté de montrer, que si on pouvait effectivement penser que Mill n'était pas opposé à une certaine pratique du paternalisme, on ne pouvait pas lui appliquer un raisonnement de type kantien. Voir..

<sup>113</sup> S. Lichtenstein, P. Slovic, *The Construction of Preference*, Cambridge: Cambridge University Press, 2006.



comportemental qui ne le serait pas, dans la mesure où – chez Kant comme chez Rousseau – une action destinée à rétablir l'autonomie de l'individu est par nature libératrice.

L'absence de fondements moraux est donc la première critique que l'on peut adresser à la théorie de Sunstein et Thaler. Plus surprenant, pour un économiste, est cependant le manque de justification économique dont fait l'objet le paternalisme comportemental. Le fait que la rationalité des agents soit compromise par un certain nombre de biais cognitifs ne justifie pas en soi une justification le paternalisme comportemental. Encore faut-il prouver que cette mesure est plus efficiente que ses alternatives, à savoir le paternalisme légal et le *self-control* ou plus généralement les stratégies visant à améliorer le contrôle de soi. Ces stratégies, déjà étudiées par Thomas Schelling<sup>114</sup>, sont aujourd'hui l'objet de la picoéconomie<sup>115</sup>. Une appréciation correcte de la pertinence de chacune de ces solutions implique donc une analyse comparative de leur désirabilité. Or, cette analyse fait défaut dans les travaux de Sunstein et Thaler. Elle est pourtant essentielle. Les résultats de l'étude comparative que j'ai proposée<sup>116</sup> prouvent, en effet, qu'aucune de ces trois solutions n'est absolument meilleure que les autres, mais que leur efficacité relative dépend conjointement de la force de volonté et du degré de confiance en soi des agents.

L'article 'Economics of paternalism : the hidden costs of self-commanding strategies' part du constat qu'il existe un certain nombre d'actions objectivement irrationnelles, qui sont à la fois désapprouvées par le spectateur impartial et par l'agent même, qui préférerait ne pas les avoir menées. C'est le cas, par exemple, des actions associées à une actualisation

---

<sup>114</sup> T.C. Schelling, "Self-Command in Practice, in Policy, and in a Theory of Rational Choice", *The American Economic Review*, 74:2, 1984, pp.1-11.

<sup>115</sup> G. Ainslie, *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person*, Cambridge: Cambridge University Press, 1992.

<sup>116</sup> C. Salvat, "Economics of paternalism: The hidden costs of self-commanding strategies", *The Journal of Philosophical Economics. Reflections on Economic and Social Issues*, IX: 1, 102-24.

hyperbolique. A la différence de l'actualisation exponentielle qui est à taux constant, et qui est communément utilisée par les économistes, l'actualisation hyperbolique est une actualisation à taux décroissant<sup>117</sup>.

En d'autres termes, l'actualisation hyperbolique est une forme de l'impatience. Plus l'évènement attendu est proche dans le temps, plus on actualise l'utilité à venir. A un certain moment, cela nous conduit logiquement à modifier nos préférences et à remettre en cause les choix opérés dans le passé, quitte à le regretter plus tard. L'actualisation hyperbolique est notamment à l'origine de la procrastination<sup>118</sup> et de la dépendance<sup>119</sup>, communément considérées comme des anomalies du choix intertemporel. La fréquence de ce type de comportement ne doit pas être sous-estimée. L'actualisation hyperbolique est en fait très répandue, et George Ainslie a montré qu'on pouvait également la rencontrer chez d'autres espèces animales, en particulier le pigeon<sup>120</sup>.

Il existe, ainsi que Strotz le montre<sup>121</sup>, deux types de comportements face à l'actualisation hyperbolique : le comportement des 'naïfs', c'est-à-dire de ceux qui ne semblent pas même s'apercevoir de la perte d'utilité occasionnée par leurs changements de préférence, et celui

---

<sup>117</sup> G. Ainslie and N. Haslam, 'Hyperbolic Discounting', in G. Loewenstein and J. Elster (eds.), *Choice over Time*, New York: Russell Sage Foundation, 1992, pp.57-92; S. Frederick, G. Loewenstein and T. O'Donoghue, 'Time Discounting and Time Preference: A Critical Review' in G. Loewenstein, D. Read and R.F. Baumeister (eds.), *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, New York: Russell Sage Foundation, 2003, pp.13-86.

<sup>118</sup> G. Akerlof, 'Procrastination and Obedience', *The American Economic Review*, 81(2), 1991, pp.1-19; T. O'Donoghue and M. Rabin, 'Doing It Now or Later', *The American Economic Review*, 89(1), 1999, pp.103-24.

<sup>119</sup> R.J. Herrnstein and D. Prelec, 'A Theory of Addiction', in G. Loewenstein and J. Elster (eds.), *Choice over Time*, New York: Russell Sage Foundation, 1992, pp.331-60; F. Gul and W. Pendorff, 'Harmful Addiction', *The Review of Economic Studies*, 74(1), 2007, pp.147-72.

<sup>120</sup> G. Ainslie, *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person*, Cambridge: Cambridge University Press, 1992.

<sup>121</sup> R.H. Strotz, 'Myopia and Inconsistency in Dynamic Utility Maximization', *The Review of Economic Studies*, 23 (3), 1955, pp.165-180.

des 'avertis' ('sophisticated') qui ont la connaissance, sinon la maîtrise, de leurs travers. Les préférences initiales de l'agent complètement naïf ne peuvent être satisfaites sans intervention paternaliste. Dans le cas des agents avertis, en revanche, une seconde alternative est possible : la mise en place de règles personnelles par l'agent lui-même. Ces règles auto-imposées peuvent être plus ou moins coercitives et peuvent parfois impliquer l'intervention d'un tiers pour exécuter la sanction. Elles comprennent, notamment, le transfert d'autorité à autrui, la suppression de certaines possibilités de choix, la mise en place d'un système de sanctions et/ou de récompenses ou encore le contrat moral<sup>122</sup>. Plus l'agent est faible de volonté (ou plus la tentation est importante), plus les règles qu'il s'impose doivent être contraignantes. Ulysse, qui sait qu'il ne pourra pas résister à l'appel des Sirènes, n'a ainsi d'autres choix que d'ordonner à ses matelots de le lier au grand mât<sup>123</sup>. En pratique, les agents ont souvent à la réprobation morale ou sociale des autres pour se contraindre à respecter leur règle. C'est sur ce schéma-là que fonctionne notamment l'association des Alcooliques Anonymes. Mais la force de volonté n'est pas le seul critère à prendre en compte. Les économistes Roland Bénabou et Jean Tirole ont ainsi mis en évidence le rôle déterminant de la confiance en soi dans le respect des règles personnelles<sup>124</sup>. Leur modèle part cependant du principe, théoriquement très discuté et empiriquement non vérifié, que plus les agents respectent leurs règles personnelles, plus ils ont intérêt à les respecter dans le futur pour protéger leur 'capital réputationnel'. Il

---

<sup>122</sup> T. Schelling, 'Self-Command in Practice, Policy, and in a Theory of Rational Choice', *The American Economic Review*, 74(2), 184, pp.1-11; T. Schelling, 'Self-Command: A New Discipline', in G. Loewenstein and J. Elster (eds.), *Choice over Time*, New York: Russell Sage Foundation, 1992, pp.167-76.

<sup>123</sup> J. Elster, *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge: Cambridge University Press, 1979.

<sup>124</sup> R. Bénabou and J. Tirole, 'Self-Confidence and Personal Motivation', *The Quarterly Journal of Economics*, 117(3), 2002, pp.871-915; R. Bénabou and J. Tirole, 'Willpower and Personal Rules', *Journal of Political Economy*, 112(4), 2004, pp.848-86.

implique une parfaite transparence et honnêteté intellectuelle des agents, qui – affirment Bénabou et Tirole – révisent périodiquement leur confiance en soi en fonction de leurs résultats.

Le modèle que j’ai proposé<sup>125</sup> reposait, au contraire, sur l’hypothèse selon laquelle les niveaux de confiance des agents sont exogènes et potentiellement déconnectés du niveau réel de leur force de volonté. Il permet de comparer les coûts respectifs du paternalisme et des règles personnelles en fonction de la force de volonté des agents et de leur niveau de confiance en soi.

Agents	A	B	C	D	E
Confiance en soi	1	$\frac{4}{5}$	$\frac{3}{5}$	$\frac{2}{5}$	$\frac{1}{5}$
Volonté	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{2}{5}$
Règle personnelle (sévérité)	Nulle	Légère	Moyenne	Appropriée	Excessive
Intervention extérieure (coût relatif de)	Fort	Moyen	Faible	Nul	Négatif

<sup>125</sup> C. Salvat, “Economics of paternalism: The hidden costs of self-commanding strategies”, *The Journal of Philosophical Economics. Reflections on Economic and Social Issues*, IX: 1, 2015, pp. 102-24.

Lorsque l'agent est parfaitement informé de son biais (comme c'est le cas de l'agent D), il met en place des règles personnelles dont la sévérité est suffisante pour le forcer à s'y plier, sans pour autant être excessive. En revanche, plus il surestime sa capacité à se conformer à ses choix intertemporels, moins ses règles personnelles sont efficaces et plus la nécessité d'une intervention extérieure se fait ressentir. Au contraire, un manque de confiance en soi peut le conduire à mettre en place des règles personnelles excessivement contraignantes, rendant le coût relatif d'une intervention paternaliste négatif.

Que peut-on conclure de ces travaux ? Tout d'abord, et c'est peut-être le point le plus significatif, le coût du paternalisme en termes d'opportunités de choix ou de liberté d'action doit être mesuré comparativement aux autres alternatives possibles. Or, dans certains cas, la mise en place de règles personnelles peut s'avérer également très contraignante pour l'agent. Le second point qu'il me semble important de mettre en avant est que le paternalisme et la règle personnelle ne sont pas des alternatives exclusives l'une de l'autre. Il est possible de montrer, au contraire, que la réponse optimale à l'actualisation hyperbolique repose sur une stratégie commune, dans laquelle l'intervention paternaliste viendrait aider l'agent à se conformer à ses propres règles. C'est d'ailleurs, me semble-t-il, à cette seule condition que les *nudges* peuvent montrer leur efficacité.

### C) THEORIE REDUCTIONNISTE DE L'IDENTITE PERSONNELLE

La question de l'identité personnelle peut, au premier abord, sembler étrangère aux préoccupations des économistes. Elle lui est pourtant centrale, et de plus en plus d'économistes en ont pris conscience au point d'être devenu un thème central de la

philosophie économique. Elle répond à un besoin des économistes, celui d'aller plus loin dans leur façon de rendre-compte du choix individuel et des interactions sociales.

Derrière la question de l'identité personnelle se profile le principe de l'égoïsme rationnel, ou plus exactement celui de la rationalité individuelle, sur lequel repose toute la théorie de la décision. S'il est démontré, par exemple, que l'identité personnelle se réduit à un certain nombre de liaisons ou de connexions psychologiques, qui s'atténuent avec le temps, un agent économique devrait logiquement préférer son présent à son lointain futur, quitte à ne pas maximiser son utilité totale sur le long terme. On peut même aller plus loin. Si l'on peut montrer, comme le fait Parfit, que l'identité personnelle n'est pas importante en soi, mais que seule compte la connexité psychologique, alors ce même agent économique peut être logiquement amené à donner la préférence à un tiers plutôt qu'à son futur soi. Un examen approfondi des fondements de l'identité personnelle peut donc remettre en cause deux hypothèses fondamentales de la science économique, celle de l'intérêt personnel et celle de la neutralité par rapport au temps.

J'ai ainsi étudié l'impact de l'approche réductionniste de l'identité personnelle sur notre conception de la rationalité dans un article paru en 2015 dans *Oeconomia* ainsi que dans l'entrée 'Parfit' de l'*Encyclopédie Philosophique*<sup>126</sup>. Je présente ci-dessous les principaux résultats de ces travaux. Je propose de retenir ici trois traits essentiels de sa pensée. Le premier est son héritage lockéen et son opposition radicale au dualisme ontologique. Le second est que, chez Parfit, ce n'est pas l'identité personnelle qui importe mais la Relation R. Je montre qu'il est possible, selon Parfit, de ne pas respecter les règles de rationalité

---

<sup>126</sup> C. Salvat, "Parfit l'égoïsme rationnel et la question de l'identité personnelle", *Oeconomia, History, Methodology, Philosophy*, 5(4), 2015, pp.437-60 ; C. Salvat, "Parfit", in *Encyclopédie Philosophique*, janvier 2019, <http://encyclo-philosophie.fr/parfit-a-2/>.

imposées par le principe de neutralité par rapport au temps sans pour autant être complètement irrationnel. Cela ne permet pas, cependant, d'éviter tous les comportements imprudents.

Il est important, tout d'abord, pour Parfit de réaffirmer, et surtout de démontrer, l'impossibilité du dualisme ontologique. La thèse dualiste de l'identité personnelle est celle notoirement défendue par Descartes bien sûr, mais plus récemment par Richard Swinburne<sup>127</sup>. Elle stipule que le sujet ne serait se réduire à son corps mais à son essence (son esprit, son âme etc) qui n'est pas matérielle. Parfit l'appelle alternativement approche simple ou non-réductionniste. Par opposition, les approches réductionnistes considèrent que l'identité peut être déterminée en termes de continuité physiques et/ou psychologiques. John Locke a beaucoup contribué à répandre l'idée que seule comptait la continuité psychologique. Ainsi, le fait que je puisse me souvenir de mes expériences passées prouve que l'individu qui a vécu ces expériences et celui qui se les remémore est le même. La version originale qu'en a donné Locke a cependant été beaucoup critiquée, par Butler et Hume notamment. La thèse de Locke présente trois faiblesses bien connues : sa circularité (Butler), le problème de la transitivité de la continuité psychologique (Reid) et la question de l'intention<sup>128</sup>.

L'objection de Butler à Locke est bien connue. Butler reproche à Locke de ne pas avoir réalisé que la notion de conscience qu'il emploie, et à partir de laquelle il infère l'identité, présuppose celle-ci. Je ne peux me souvenir que de *mes* expériences passées donc ma

---

<sup>127</sup> S. Shoemaker and R. Swinburne, *Personal Identity. The Dualist Theory*, Oxford: Basic Blackwell, 1984.

<sup>128</sup> D. Shoemaker, "Personal Identity and Ethics" In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2014. <<http://plato.stanford.edu/archives/spr2014/entries/identity-ethics/>>.

mémoire révèle une relation d'identité qui lui préexiste. Elle ne peut donc l'établir. Parfit répond à l'objection de Butler grâce à la notion de quasi-mémoire, introduite par Shoemaker<sup>129</sup>. Contrairement à Butler, qui suppose que la mémoire reproduit nécessairement une expérience passée de celui qui l'a vécue, Shoemaker et Parfit conçoivent la (quasi-)mémoire comme la reproduction virtuelle d'évènements passés, mais non nécessairement vécus par la personne qui s'en souvient. Ainsi, pour Parfit, un souvenir ordinaire n'est qu'une forme particulière du quasi-souvenir. Il se distingue simplement par le fait que c'est *moi* qui ait vécu les évènements dont je me souviens. Dans ce cas-là, l'identité personnelle ne présuppose pas le souvenir, et la continuité psychologique (ou autrement dit ici, la quasi-mémoire) peut tout à fait être l'un des éléments constitutifs de l'identité.

Seconde objection possible à la version lockéenne de l'approche psychologique : la perte de mémoire ne constitue pas nécessairement une perte d'identité. Thomas Reid est le premier à avoir soulevé le problème. Il prend l'exemple suivant : un officier s'empare d'un étendard ennemi sur un champ de bataille, il se souvient avoir fouetté lorsqu'il était petit pour avoir volé des pommes dans un verger. L'officier courageux (B) est donc bien le même que l'enfant voleur (A). Vers la fin de sa vie, l'officier, devenu général (C), se souvient avoir pris l'étendard aux ennemis mais ne se souvient pas avoir été puni pour vol lorsqu'il était petit. Selon la théorie de Locke, l'officier et le général sont une et même personne, en revanche, le général n'est plus la même personne que l'enfant qui volait des fruits sur les arbres.

---

<sup>129</sup> S. Shoemaker, "Persons and Their Pasts", *American Philosophical Quarterly*, 7, 1970, pp.269–285.



C=B, B=A mais C≠A

Pour être acceptable, la version lockéenne de l'approche psychologique doit donc être révisée. Une première modification consiste à envisager la continuité psychologique comme une *série* indirecte de souvenirs qui se recourent dans le temps. *Le fait que je me souviens d'avoir été l'officier, qui lui-même se souvenait d'avoir été l'enfant puni, suffit à me prouver que j'ai bien été celui qui fut fouetté pour avoir volé une pomme.* La transitivité de la relation d'identité est ainsi rétablie. Mais cette modification ne se fait-elle pas au détriment de l'approche *psychologique* de la notion d'identité personnelle ? Permet-elle de rendre compte de l'évolution ou des changements identitaires des individus ? Tout l'intérêt de la théorie de Locke était, rappelons-le, qu'elle permettait de reconnaître aux individus la possibilité d'être différents au cours de leur vie. La continuité psychologique est un marqueur identitaire, elle relate un état de proximité psychologique, qui lui-même atteste d'une relation d'identité. Le tort de Locke est certainement de l'avoir réduit à la seule mémoire. Dans son esprit, si le vieux général ne se souvient plus d'avoir été puni, c'est parce qu'il n'est psychologiquement plus la même personne. Donc s'il est vrai que

« *Le fait que je me souviens d'avoir été l'officier, qui lui-même se souvenait d'avoir été l'enfant puni, suffit à me prouver que j'ai bien été celui qui fut fouetté pour avoir volé une pomme.* »

Il est incorrect d'en conclure que

« *Le fait que je me souviens d'avoir été l'officier, qui lui-même se souvenait d'avoir été l'enfant puni, suffit à me prouver que suis toujours celui qui fut fouetté pour avoir volé une pomme.* »

La simple continuité psychologique ne suffit pas à établir l'identité des personnes dans le temps. Celle-ci doit également reposer, nous dit Parfit, sur un degré minimum de *connexité psychologique* (*psychological connectedness*). La connexité psychologique peut être définie comme la relation directe, plus ou moins forte, qu'il existe entre deux états psychologiques. A la différence de la simple continuité, la connexité psychologique admet des degrés, et donne ainsi la profondeur qui manquait chez Locke à la notion de continuité.

La continuité psychologique est une relation transitive qui établit que je suis lié à telle ou telle partie de mon futur en raison de toutes les parties intermédiaires de mon futur avec lesquelles je suis déjà lié. Ainsi dans la séquence [i1-i10], je peux dire que la personne i en temps 1 est liée à la personne i en temps 10 parce que cette dernière est psychologiquement proche de i9, qui est elle-même est proche de i8 etc... jusqu'à i1. C'est l'identité lockéenne. L'identité se fonde sur la mémoire : je sais que i10 est la même personne que i1 parce que i10 se souvient d'avoir été i9, qui se souvient d'avoir été i8 etc.. Mais il est également possible de considérer séparément les éléments de cette séquence. C'est ce que Parfit appelle la connexité psychologique. Ainsi je peux m'interroger sur la relation entre i1 et i6 et entre i3 et i10 indépendamment l'une de l'autre. Il n'y a plus nécessairement de transitivité entre mes mois temporels, et ce n'est plus la même chose de se demander s'il y a identité entre i1 et i10, d'une part, et i3 et i6, d'autre part. Ce qui compte maintenant c'est la relation *directe* qu'entretient i1 avec i6 ou i3 avec i10. Or certains éléments de cette relation tiennent à différents degrés dans le temps. Bien que les séquences (i1-i10) et (i3-i6) soient *également* continues, il n'y a pas nécessairement le *même* degré de connexité entre i1 et i10 qu'entre i3 et i6.

L'existence de quelques souvenirs qui se recourent dans le temps ne suffit pas à établir une relation d'identité. En revanche, l'existence de fortes proximités psychologiques, qui se succèdent et se recourent dans le temps, permet d'envisager une telle relation. Naturellement, et j'aurais l'occasion d'y revenir ci-dessous, cette relation d'identité est d'une nature différente de celle initialement conçue. Chez Locke, comme chez la plupart des réductionnistes (et *a fortiori* des non-réductionnistes), l'identité est déterminée par l'existence d'un lien psychologique continu. Dans sa version stricte, ce lien est direct mais on a vu qu'il pouvait également être indirect. Dès lors que l'existence de ce lien est établie, l'identité dans le temps est assurée. L'identité personnelle est donc '*tout ou rien*' ('*all-or-nothing*'). La notion de connexité, en revanche, repose sur une évaluation qualitative du lien psychologique (qui, lui, est toujours direct). Pour Parfit, on ne peut parler de continuité psychologique que s'il existe une série ou une chaîne de *fortes* connexités, qui se chevauchent dans le temps. Rien, cependant, ne nous permet de déterminer avec précision une forte d'une faible connexité. Lorsque les connexions directes entre deux états psychologiques sont *suffisamment* nombreuses, on peut parler – nous dit Parfit – de forte connexité<sup>130</sup>.

La troisième objection soulevée contre Locke est que la mémoire ne peut raisonnablement constituer le seul élément pertinent pour juger de l'existence d'un lien psychologique. Ainsi qu'on a déjà pu le dire, l'amnésie entraîne automatiquement, chez Locke, la perte d'identité. C'est excessif. Notre rapport à nous-même ne repose pas exclusivement sur la conscience que l'on a de notre passé. Elle inclut également, entre autres, le rapport que l'on entretient avec notre futur. Le fait que l'on partage, à différentes époques de notre

---

<sup>130</sup> D. Parfit, *Reasons and Persons*, *op. cit.*, p.206.

vie, les mêmes intentions est un élément aussi digne d'être considéré pour déterminer notre identité que celui que l'on partage, durant ces mêmes périodes, les mêmes souvenirs. Ce point est d'ailleurs généralement admis aujourd'hui. Le danger, nous rappelle Parfit, est qu'en ajoutant simplement l'intention à la mémoire, on risque de retomber dans le même piège que celui dans lequel était tombé Locke. Pour éviter tout risque de circularité, Parfit propose donc d'utiliser le concept de *quasi-intention*<sup>131</sup>. Une personne forme une quasi-intention lorsqu'elle se destine à accomplir les intentions d'une autre personne. Dans certains cas, mais dans certains cas seulement, cette autre personne n'est autre qu'elle-même. Ainsi dans le cas de l'intention, comme dans le cas de la mémoire, les connexions psychologiques ne présupposent pas l'identité personnelle.

La question de l'indétermination de l'identité personnelle se pose lorsqu'on considère des cas extrêmes voire des cas imaginaires improbables. Ces exemples fictionnels, souvent tirés de la science-fiction, sont importants car ils nous permettent de sonder nos intuitions, dont certaines sont infondées. C'est notamment le cas, pour Parfit, de l'identité personnelle. Les approches néo-lockéennes considèrent que l'identité personnelle repose exclusivement sur la continuité (et, dans le cas de Parfit, la connexité) psychologique. Lorsque celle-ci est attestée, l'identité est déterminée. Dans l'exemple du prince et du cordonnier de Locke, dans lequel la mémoire du prince est transférée dans le cerveau du cordonnier (qui lui ne garde aucun souvenir de sa précédente vie), aucun doute n'est permis : c'est le prince qui se réveille dans le corps du cordonnier. De façon similaire, dans le cas de la téléportation, c'est bien la même personne qui renaît sur Mars, dans un corps certes absolument

---

<sup>131</sup> D. Parfit, *Reasons and Persons*, op. cit., p.261.

identique au sien, mais néanmoins dans un autre corps. Mais que se passerait-il si la mémoire (ou plus généralement l'état psychologique) d'un individu était transférée non pas à un autre mais à plusieurs autres individus ? Imaginons, nous dit Parfit – qui s'inspire ici des travaux de Wiggins<sup>132</sup> –, que mon corps et que les cerveaux de mes deux frères jumeaux soient irrémédiablement détruits dans un accident. Mon cerveau est séparé en deux et chaque moitié est transplantée avec succès dans chacun des corps de mes deux frères. Chacun croit être moi, semble se remémorer ma vie, possède ma personnalité et est psychologiquement continu avec moi. Etant issu de triplés, chacun a par ailleurs un corps identique au mien<sup>133</sup>. L'un et l'autre peuvent-ils à la fois revendiquer le fait d'être moi ?<sup>134</sup>

Quatre configurations sont possibles : (1) je ne survis pas, (2) je survis en tant que l'un de mes frères, (3) je survis en tant que l'autre, (4) je survis en tant que l'un et l'autre<sup>135</sup>. Parfit refuse la première possibilité. Il est possible, nous dit-il, de survivre avec une partie seulement de son cerveau. On peut donc théoriquement envisager le fait qu'une seule partie de mon cerveau soit transplantée et que l'autre soit détruite. Je survivrai alors dans le corps de la personne dans laquelle on a transplanté la moitié de mon cerveau. Mais si je peux survivre avec la moitié seulement de mon cerveau (transplanté), pourquoi ne survivrais-je pas dans le cas où la deuxième partie de mon cerveau est également transplantée au lieu d'être détruite ? Il s'agit là de deux événements indépendants l'un de l'autre. Ma survie dans un nouveau corps, avec la moitié de mon cerveau, ne dépend pas

---

<sup>132</sup> D. Wiggins, *Identity and spatio-temporal continuity*, London: Blackwell, 1967.

<sup>133</sup> D. Parfit, *Reasons and Persons*, *op. cit.*, pp.254-55.

<sup>134</sup> Un cas similaire se pose dans le cas de la télétransportation dans lequel Parfit envisage, en première instance, la destruction de mon corps sur Terre, et dans un deuxième temps, me fait communiquer avec mon clone sur Mars, D. Parfit, *Reasons and Persons*, *op. cit.*, pp.199-200.

<sup>135</sup> D. Parfit, *Reasons and Persons*, *op. cit.*, p.256.

du sort réservé à l'autre moitié de mon cerveau. De la même manière, si chaque moitié de mon cerveau est transplanté avec succès dans le crâne de l'un et de l'autre de mes frères jumeaux, pourquoi survivrais-je dans l'un plutôt que dans l'autre ? Les deux opérations étant strictement identiques et indépendantes, elles ne peuvent avoir des résultats opposés. Ne reste donc que la dernière possibilité, celle de ma survie dans l'un et l'autre corps de mes frères.

Cela peut surprendre. Notre réaction première est de refuser le fait qu'on puisse être à la fois l'un et l'autre. L'identité, telle que nous la concevons implique l'unicité, et il serait absurde d'imaginer que je puisse dialoguer avec moi-même. Certes, mais qui nous dit que nos intuitions sont vraies ? Ne confondons-nous pas identité numérique et identité qualitative (§76) ? Pourquoi accepterions-nous le fait que le Prince puisse se réveiller dans le corps du cordonnier mais pas le fait que je puisse me réveiller dans le corps de mes deux frères ? Mais, plus fondamentalement, peut-être ne nous posons-nous pas les bonnes questions. L'important, nous dit Parfit, n'est pas vraiment de savoir si je suis l'un de mes frères, les deux ou aucun des deux, mais de savoir si je continue à exister à travers eux. Parfit n'affirme pas que ce qui fait l'identité d'une personne puisse être reproduit à volonté (pensez à l'exemple de la télétransportation), il montre seulement que la continuité psychologique sur laquelle repose le concept d'identité n'a pas nécessairement lieu au sein du même individu, mais qu'elle peut *théoriquement* avoir lieu entre deux individus, ou plus. Le fait de savoir si je *serai* l'un ou l'autre de mes frères, ou les deux, est une question dénuée de sens (*'an empty question'*).

La vraie question n'est donc pas de savoir quelle est la meilleure description du phénomène observé mais de savoir en quoi ce phénomène nous affecte. Certaines personnes

considèrent la division comme aussi mauvaise, ou presque aussi mauvaise, que la mort. Cette réaction, affirme Parfit, est irrationnelle<sup>136</sup>. Parce qu'elle assure la continuité psychologique de l'individu, la division devrait être considérée, au contraire, comme aussi bonne que la survie ordinaire. Le fait qu'il existe entre moi et une autre entité une forte continuité et connexité psychologique, ce que Parfit appelle une *Relation R*, est plus important que le fait de savoir s'il s'agit encore de moi ou non. Parfit peut donc affirmer que l'identité personnelle n'est pas ce qui importe. Ce qui est compte vraiment est la Relation R, c'est-à-dire la connexité et/ou continuité psychologique, Relation qui peut soit être le fait d'une cause normale (dans le cas d'une vie ordinaire) soit le fait de n'importe quel type de cause (dans le cas de la télétransportation ou de la greffe de cerveau par exemple). Dans la plupart des cas, les cas normaux, la Relation R recouvre la notion d'identité personnelle. Parfit s'appuie donc, dans un premier temps, sur le cas de R avec le type normal de cause ('with the right kind of cause')<sup>137</sup>. Il la réintègre ensuite dans R comme étant simplement un cas parmi d'autres. On peut ainsi avoir l'impression que la question de l'identité est fondamentale. Elle ne l'est, dans ce cas, que par coïncidence.

Nous avons tendance à penser qu'il ne peut y avoir qu'une seule façon de répondre à la question 'Est-ce que je vais mourir?', soit par 'oui' soit par 'non'. De la même manière, nous avons tendance à penser que notre identité est sans équivoque. Ces réactions révèlent notre non-réductionnisme instinctif. Le réductionnisme nous enseigne, nous dit Parfit, que l'identité personnelle n'est pas déterminée et que, par conséquent, la question de notre survie est moins catégorique qu'il n'y paraît. Puisque la thèse du réductionnisme semble acquise, alors il semble logique pour Parfit que nous devions abandonner nos

---

<sup>136</sup> D. Parfit, *Reasons and Persons*, op. cit., p.261.

<sup>137</sup> D. Parfit, *Reasons and Persons*, op. cit., p.215.

certitudes identitaires, et accepter le fait qu'on ne puisse pas répondre à des questions telles que 'Est-ce que je vais mourir ?'. Cela est, cependant, aller un peu vite en besogne et certaines résistances se manifestent au sein même du réductionnisme.

La thèse de l'indétermination de l'identité a deux implications théoriques importantes. Tout d'abord, elle implique – ou peut impliquer – que les relations psychologiques et physiques qui forment ce que l'on désigne généralement par le terme d'identité peuvent en fait être décrites sans référence à la personne. « Bien que les personnes existent », écrit Parfit, « il est possible de donner une description *complète* de la réalité *sans* prétendre que les personnes existent »<sup>138</sup>. Cette affirmation a suscité beaucoup de commentaires et de confusions, obligeant Parfit à préciser qu'il ne défendait en aucun cas une version impersonnelle de l'identité<sup>139</sup>.

La deuxième conséquence théorique du réductionnisme (ou de l'indétermination puisque l'une ne va pas sans l'autre chez Parfit) est que l'identité personnelle n'a que très peu d'importance morale et/ou rationnelle. La question de l'importance morale de l'identité personnelle renvoie notamment aux débats portants sur la séparabilité des personnes et la justice distributive. Elle ne sera pas traitée ici. Celle de l'importance rationnelle de l'identité porte sur ce qu'on appelle communément la prudence, c'est-à-dire le rapport que l'on entretient vis-à-vis de notre futur bien-être. C'est l'un des fondements du dogme de l'égoïsme rationnel, qui considère que la rationalité individuelle consiste à se préférer aux autres mais à ne pas faire de préférence entre son bien-être présent et son bien-être futur. Mais, le principe prudentiel ne peut se justifier qu'à condition que je sois la même personne

---

<sup>138</sup> D. Parfit, *Reasons and Persons*, *op. cit.*, p.212.

<sup>139</sup> Dans le sens où les expériences n'ont pas de sujets, et les pensées n'ont pas de penseurs.



aujourd'hui et dans le futur (sinon je viole le premier postulat puisque je ne me préfère pas aux autres).

L'implication pratique de cette thèse est énorme. Cela change, ou peut potentiellement changer, notre rapport à notre futur. Pour un non-réductionniste, comme Swinburne, il est évident que la seule raison pour laquelle nous nous soucions de notre futur, est précisément qu'il s'agit de *notre* futur. Si, maintenant, on accepte la thèse réductionniste de Parfit, il ne s'agit plus à proprement parler de *notre* futur mais du futur d'une personne avec laquelle nous entretenons une continuité (et/ou connexité) psychologique plus ou moins grande. Il est alors possible d'affirmer que, si l'on adopte une position réductionniste, alors on a plus de raison de se soucier de son futur. C'est ce que Parfit appelle la *Proposition Extrême* ('*The Extreme Claim*'), et dont il attribue la première formulation à Butler<sup>140</sup>.

Il est important de noter, tout d'abord, que Parfit avait initialement rejeté cette proposition. Dans son article 'Personal Identity and Rationality' (1982), Parfit avait tout d'abord soutenu l'idée que les non-réductionnistes avaient tort de croire que le fondement du souci de soi était dans ce quelque chose de plus qui constituait selon eux l'identité personnelle. Dans une métaphore très inspirée, il comparait alors ce 'deep further fact' au Soleil qui éblouit et cache la présence constante, mais du coup moins remarquée, de la Lune, censée représenter la continuité psychologique<sup>141</sup>. Sensible aux critiques de Wachsberg, Parfit revient sur sa position dans *Reasons and Persons*, et admet que d'un point de vue non-réductionniste, il est légitime de penser – qu'en devenant réductionniste

---

<sup>140</sup> D. Parfit, *Reasons and Persons*, *op. cit.*, p.307.

<sup>141</sup> D. Parfit,, «Personal Identity and Rationality», *Synthese*, vol. 53, n°2, 1982, p.230.

– nous perdriions toute raison de nous préoccuper de notre futur<sup>142</sup>. D'un autre côté, les non-réductionnistes peuvent arguer du fait qu'on se préoccupe tout aussi bien, et parfois mieux, du futur des personnes qui nous sont chères que de notre propre personne, et que ce n'est donc pas l'identité personnelle mais la Relation R qui compte dans notre relation au futur. C'est ce que Parfit appelle la *Proposition Modérée* (*The Moderate Claim*). L'argument est un peu troublant. Si, en se convertissant au réductionnisme, un non-réductionniste pense devoir perdre toute préoccupation pour son futur, c'est qu'il continue à penser comme un non-réductionniste.

Comment répondre à l'objection de la proposition extrême ? Le premier argument de Parfit est que les hommes sont, dans leur très grande majorité, instinctivement non-réductionnistes. Nous avons instinctivement peur de la mort, si ce n'était pas le cas l'espèce humaine serait depuis longtemps éteinte<sup>143</sup>. Nous pouvons nous convaincre que cette peur n'est pas rationnelle, qu'elle n'est pas théoriquement justifiée, elle continuera néanmoins à nous effrayer. A tout le moins, notre conception de l'identité peut-elle dédramatiser la perspective de notre propre mort, mais elle ne pourra jamais nous rendre totalement indifférents à son égard. Nous sommes naturellement conditionnés à craindre notre mort, et par conséquent à éviter toutes les actions qui pourraient la précipiter. Cela constitue une première réponse possible à la Proposition Extrême.

Une seconde possibilité est de voir dans la prudence une norme morale plutôt qu'une norme de rationalité. Parfit suggère trois justifications morales possibles de la prudence<sup>144</sup>.

---

<sup>142</sup> L'argument est un peu troublant. Si, en se convertissant au réductionnisme, un non-réductionniste pense devoir perdre toute préoccupation pour son futur, c'est qu'il continue à penser comme un non-réductionniste.

<sup>143</sup> D. Parfit, *Reasons and Persons*, *op. cit.*, p.308.

<sup>144</sup> D. Parfit, *Reasons and Persons*, *op. cit.*, p.106.

La première est d'inspiration utilitariste, la deuxième d'inspiration kantienne et la troisième, disons, d'inspiration parfitienne. Du point de vue utilitariste, les imprudences sont *moralement* condamnables dans la mesure où elles amoindrissent la somme totale des utilités. Or, à moins que l'imprudence d'une personne profite aux autres, c'est presque toujours le cas. En plus les désutilités personnelles qu'elle génère, l'imprudence se caractérise par de fortes externalités négatives. L'imprudence d'un individu a généralement un coût affectif et/ou financier pour les proches mais aussi pour la collectivité qui doit souvent en payer le prix (assistance médicale et/ou sociale, primes d'assurances, surcroît de réglementation etc..). D'un point de vue kantien ou plutôt néo-kantien, je ne dois pas faire de différence entre moi et les autres. On notera le retournement ironique opéré ici par Parfit. Alors que l'argument kantien est destiné à convaincre du bien-fondé de l'altruisme et du sacrifice de soi, il est utilisé ici pour justifier des actions égoïstes. Le raisonnement est le suivant : s'il n'y a pas de différence entre moi et les autres, et que je me dois aux autres, alors je me dois également à moi-même. Certes, je ne me dois pas plus à moi-même qu'aux autres, mais suffisamment pour ne pas me mettre dans des situations difficiles<sup>145</sup>.

La troisième justification morale de la prudence est la suivante. S'il n'existe pas ou peu de connexité psychologique entre mon moi présent et mon moi futur, alors je dois me comporter envers celui-ci comme je me comporterais envers quelqu'un qui n'existe pas encore. Autrement dit, d'un point de vue moral, il n'est pas injustifié d'affirmer que les

---

<sup>145</sup> L'argument mériterait en fait d'être approfondi car il repose implicitement sur le postulat que le paternalisme est toujours légitime. Je suis en droit d'intervenir pour mon bien-être futur car je suis en droit d'intervenir pour le bien-être futur des autres, et je ne suis pas différent d'eux. Mais le paternalisme n'est pas justifiable, ou l'est difficilement, chez Kant. Il y a une différence entre agir comme si je n'étais pas différent des autres et agir pour moi comme j'agirais pour les autres. Or, Parfit n'élabore pas sur cette différence.

personnes-dans-le-futur ont le même statut que les générations à venir. Donc, si nous partons du principe que nous avons un certain devoir moral envers les générations à venir, alors nous avons également un devoir moral envers notre propre futur. Le problème est que Parfit ne parvient pas à démontrer, dans la quatrième et dernière partie de *Reasons and Persons*, l'existence d'un devoir moral des générations présentes envers les générations futures.

Le problème est qu'il ait ici recours à un argument moral. Jusqu'à présent, Parfit a, en effet, été très attentif à ne pas mélanger les arguments d'ordre moral et ceux d'ordre logique, et à discuter la théorie de la rationalité *S* sur le strict plan de la rationalité. Il aurait, tout aussi bien, pu utiliser l'argument moral, qu'il considère d'ailleurs être le meilleur, dès le départ. Il s'y est refusé. Pourquoi, alors, l'utiliser à la fin, et compromettre par-là l'édifice déjà élevé ? Quant à l'argument évolutionniste, sa principale faiblesse est d'être *ad hoc*. Alors que l'argument moral tente de trouver des justifications possibles à la prudence, l'argument évolutionniste fait état d'une tendance à la prudence sans pour autant tenter de lui donner une justification. Or, et c'est ce qui – de mon point de vue – rend l'argument moral encore plus critiquable, l'argument évolutionniste suffisait à contrer la Proposition Extrême. Si, en effet, les individus luttent instinctivement pour leur survie, y compris lorsque le degré de connexité psychologique avec leur futur ne le justifie pas, alors il n'y a pas de raison que cette réaction prudentielle ne tienne pas à un degré moindre pour toutes les actions qui menaceraient la bonne conservation future de l'individu. Il est donc possible d'affirmer que le réductionnisme, bien que pouvant justifier des actions contraires, ne conduit pas nécessairement les individus à commettre de 'sérieuses imprudences'.



# TROISIEME PARTIE

## LES TROIS NIVEAUX DE L'ETHIQUE

## 1. Ethique normative et utilitarisme

Le paternalisme comportemental m'a d'abord permis de m'interroger sur les conditions de l'efficacité d'une règle pour accroître la rationalité de nos actions. Aujourd'hui, je m'interroge plutôt sur les conditions de l'efficacité d'une règle pour accroître la moralité de nos actions. La philosophie utilitariste, et sa critique déontologique, offrent un cadre théorique unique pour penser les conditions de la décision morale.

L'utilitarisme est une philosophie morale née au dix-huitième siècle sous l'impulsion d'un intellectuel anglais, Jeremy Bentham (1748-1832), qui souhaitait donner des fondements rationnels à la morale et à la justice. Bentham entendait ainsi s'attaquer à la fois à la tradition et à la jurisprudence et à l'intuitionnisme comme principes d'évaluation morale, standards qu'il considérait arbitraires et iniques.

La philosophie utilitariste classique repose sur quatre grands principes fondateurs : le conséquentialisme, c'est-à-dire le principe qu'une action est moralement bonne (*right*) si ses conséquences sont avantageuses pour l'ensemble de la communauté, l'hédonisme, c'est-à-dire l'idée selon laquelle une action est considérée comme avantageuse si elle augmente le plaisir (net) des individus, l'impartialité, c'est-à-dire l'idée selon laquelle tous les intérêts des individus doivent être traités également, et le maximalisme, soit l'idée qu'une fois le critère d'évaluation morale connu (i.e. le plaisir total), les agents sont tenus d'agir de façon à en maximiser la quantité. Ceux-ci ne sont pas, cependant, toujours caractéristiques de l'utilitarisme contemporain. Enfin, précisons que dans sa version traditionnelle, la théorie utilitariste est une procédure d'évaluation *directe* des actions *individuelles*. Dans un certain nombre de ses versions contemporaines, elle se présente, en revanche, comme une théorie d'évaluation *indirecte* des actions *individuelles* ou comme

une théorie d'évaluation *directe* des actions *collectives*. C'est cet aspect de la théorie utilitariste contemporaine qui, à mon sens est appelé à être le plus discuté dans les prochaines années au regard des enjeux posés par l'éthique appliquée.

#### A) LE PRINCIPE D'IMPARTIALITE

Le principe d'impartialité désigne le principe selon lequel une règle morale ne peut favoriser les intérêts de certaines personnes ou de certains groupes de personnes au détriment des autres. On définit alors l'impartialité comme *l'impossibilité de justifier un choix par les intérêts d'une ou de plusieurs personnes*. Nous aurons l'occasion ci-après de le distinguer de l'impersonnalisme que l'on définit comme *l'impossibilité de justifier un choix du point de vue d'une ou de plusieurs personnes*.

La notion d'impartialité est étroitement liée à deux autres notions, également très importantes dans le raisonnement utilitariste, celle de l'universalité et celle de l'égalité. Dans un certain nombre de cas, elles ne sont pas vraiment distinguées. Elles peuvent, pourtant, recouvrir différentes réalités. Le principe d'universalité, que l'on trouve déjà chez Kant, affirme que, pour qu'elle puisse être qualifiée de morale, une action doit pouvoir être universalisable, c'est-à-dire faire l'objet d'une loi universelle, s'appliquant à tous et dans toutes les circonstances. En éthique, on définit donc l'universalité à *l'impossibilité de justifier un choix à partir de situations particulières*.

A part quelques rares philosophes, dont Parfit, les utilitaristes n'ont pas recours à la notion d'universalité. Les utilitaristes de la règle lui préfèrent celle de généralité. On dit qu'une action peut être qualifiée de morale si (en plus des critères de maximisation qui la caractérisent), celle-ci est généralisable, c'est-à-dire qu'elle n'est particulière ni à une



situation ni à une personne donnée. Elle peut, néanmoins, être particulière à un type de situations ou de personnes. A la différence de l'universalisation, le processus de généralisation ne garantit pas l'impartialité des choix.

En pratique, l'impartialité consiste à évaluer les conséquences d'une action 'du point de vue de l'univers', selon l'expression de Sidgwick, ou encore, à se mettre à la place de toutes les personnes potentiellement affectées par ladite action. Autrement dit, soit on ne prend parti pour aucune personne, soit on prend parti pour toutes à la fois. Il existe donc un lien logique entre impartialité et égalité. Peter Singer considère ainsi que l'adoption de ce principe le conduit directement à adopter celui d'égalité de considération des intérêts individuels<sup>146</sup>. Nous verrons, néanmoins ci-dessous, que le fait de ne pas vouloir prendre parti ne signifie pas nécessairement traiter tous les individus de façon égale.

L'impartialité n'est pas spécifique à l'utilitarisme. Un certain nombre de philosophes pensent d'ailleurs qu'elle est indissociable de tout raisonnement moral<sup>147</sup>. Le contractualisme est typiquement une théorie fondée sur l'impartialité puisque les règles morales sont consenties par des individus alors qu'ils sont dépourvus de toute caractéristique personnelle, sociale ou pouvant être socialement pertinentes (sexe, ethnie, éducation, richesse etc.). C'est ce que l'on appelle le voile d'ignorance.

Si le principe d'impartialité n'est pas spécifique à l'utilitarisme, en revanche, la manière dont il se manifeste dans ses prescriptions l'est, et a été l'objet de fortes controverses. Associé au principe de maximisation d'utilité, l'impartialité implique, en effet, le fait de ne pas prendre en compte l'identité des agents impactés mais uniquement la quantité d'utilité

---

<sup>146</sup> P. Singer, *Practical Ethics*, Cambridge: Cambridge University Press, 1993.

<sup>147</sup> P. Singer, *Practical Ethics*, Cambridge: Cambridge University Press, 1993, p.11.

produite (ou détruite) par une action. William Godwin (1756-1836), l'un des premiers représentants de l'utilitarisme, a donné un exemple devenu célèbre de ce principe. Il imagine un incendie dans la maison de Fénélon (1695-1715), dans laquelle l'auteur des *Aventures de Télémaque* (1699) est prisonnier des flammes avec sa femme de chambre<sup>148</sup>. Vous passez devant la maison en feu mais vous ne pouvez sauver qu'une seule personne. Le principe d'utilité exige que vous sauviez Fénélon. Imaginez maintenant que la femme de chambre soit votre mère. Cette information doit-elle influencer votre décision ? Non, répond, Godwin. L'action morale doit pouvoir être impartialement considérée comme telle. Le résultat de la délibération morale ne peut dépendre de l'identité ni des intérêts personnels du passant.

S'il est aisé à justifier rationnellement, le principe d'impartialité est, cependant, souvent intuitivement difficile à défendre et pose donc la question de la valeur morale des intuitions. Le message de Godwin est qu'on a tort de se fier à nos intuitions (ici sauver sa propre mère). C'est également celui de Peter Singer, qui défend la primauté du principe d'impartialité dans la délibération morale sur celui de la proximité biologique, psychologique ou spatiale<sup>149</sup>. Selon Singer, l'intuition morale selon laquelle nous aurions une obligation morale de porter de l'aide à un enfant qui se noie devant nous mais pas à celui qui meurt de faim au Bangladesh, à un enfant auquel on est sentimentalement ou biologiquement lié mais pas à celui qui nous est inconnu, est injustifiable<sup>150</sup>. La souffrance ou la mort d'un enfant, quel qu'il soit, est dramatique et doit être évitée si elle peut l'être.

---

<sup>148</sup> W. Godwin, *An Enquiry Concerning Political Justice, and its Influence on General Virtue and Happiness*, London: G.G.J. and J. Robinson, 1793.

<sup>149</sup> P. Singer, *Practical Ethics*, Cambridge: Cambridge University Press, 1993.

<sup>150</sup> P. Singer, *Practical Ethics*, Cambridge: Cambridge University Press, 1993.

L'argument est naturellement valable pour tous les êtres humains, mais également pour Singer, on le verra, pour les animaux.

Si l'impartialité est généralement présentée comme une condition de la morale, elle peut être critiquée pour son coût jugé parfois exorbitant pour l'individu. En 1973, le philosophe britannique Bernard Williams publie 'Une critique de l'utilitarisme' dans laquelle il accuse l'utilitarisme de ne pas respecter l'intégrité des personnes<sup>151</sup>. Une action n'a de valeur que par son impact net, quelle que soit la personne qui l'entreprend, les raisons pour les lesquelles elle est entreprise et les moyens qu'elle utilise. L'impartialité qu'exige l'utilitarisme de la part des individus les empêche d'entretenir des relations humaines étroites – et nécessairement privilégiées – avec leurs amis ou leur famille. Elle rend *de facto* impossible toute vie 'ordinaire' et promeut un système dans lequel toutes nos actions sont dictées par un calcul de maximisation. L'erreur des utilitaristes, affirme également Peter Railton<sup>152</sup>, est d'avoir ignoré la dimension profondément 'aliénante' de leur philosophie morale pour les individus, au point de rendre leur comportement complètement impersonnel et comme détaché de tout attachement sentimental et social.

La critique de Williams et Railton fait écho à un autre argument avancé par le philosophe américain John Rawls dans sa *Theory of Justice*<sup>153</sup>. Rawls y accuse l'utilitarisme d'avoir confondu impartialité et impersonnalisme. La procédure morale qui garantit que les intérêts de chaque personne soient également considérés n'implique pas, pour les contractualistes, que le point de vue de la personne soit totalement éliminé. En d'autres

---

<sup>151</sup> B. Williams, "A Critique of Utilitarianism", in Smart J.J.C et Williams B., *Utilitarianism: For and Against*, Cambridge: Cambridge University Press, 1973, pp.75-150.

<sup>152</sup>P. Railton, "Alienation, consequentialism, and the demands of morality", *Philosophy and Public Affairs*, vol.13, n°2, 1984, pp.134-171.

<sup>153</sup> J. Rawls, *Théorie de la justice*, Traduit par C. Audard, Paris: Editions du Seuil, 1987 [1971].

termes, l'objectivité morale ne se réduit pas à une agrégation d'unités d'utilité isolées de la personne. Rawls récuse le raisonnement qui conduit à penser la société à l'image de l'individu et, ainsi, à justifier le sacrifice interpersonnel d'utilités. Au niveau individuel, il est possible – et dans certains cas souhaitable – de sacrifier l'utilité d'une période pour le bénéfice d'une autre ou de plusieurs autres. C'est ce prescrit la règle de la prudence, qui repose elle-même sur le principe de neutralité par rapport au temps<sup>154</sup>. En redistribuant l'utilité entre les différentes périodes de sa vie, l'agent maximise son utilité totale. L'erreur de l'utilitarisme, nous dit Rawls, est de croire (et surtout de nous faire croire) que le même raisonnement puisse également s'appliquer entre les individus, et qu'une redistribution des utilités entre les agents est objectivement souhaitable. Ce faisant, il nie la spécificité et la séparabilité de la personne<sup>155</sup>, et fait de l'utilitarisme une théorie, non pas impartiale, mais impersonnelle.

L'impartialité désigne, on l'a vu, l'impossibilité de justifier un choix moral par les intérêts d'une ou de plusieurs personnes. L'impersonnalisme désigne, quant à lui, l'impossibilité de justifier un choix moral du point de vue d'une ou de plusieurs personnes. Une approche impersonnelle de l'éthique est donc une approche dans laquelle le concept de personne est superfétatoire. Les implications morales de ce transfert sont énormes pour les individus. Au-delà du sentiment d'aliénation qu'il peut provoquer, il peut également légitimer un certain nombre de politiques de redistributions contestables ou d'atteintes à l'intégrité des personnes.

---

<sup>154</sup> D.O. Brink, 'Prospects for Temporal Neutrality', in C. Callender, *The Oxford Handbook of Philosophy of Time*, Oxford and New York: Oxford University Press, 2011, pp. 353-81.

<sup>155</sup> J. Rawls, *Théorie de la justice*, Traduit par C. Audard, Paris: Editions du Seuil, 1987 [1971], p.53.

Face aux difficultés soulevées par le principe d'impartialité, un certain nombre d'auteurs ont tenté de réintégrer la notion de personne dans le raisonnement conséquentialiste (voire utilitariste) en assouplissant le principe d'impartialité et en introduisant le concept de prérogative personnelle.

Les objections faites au principe d'impartialité sont fortes et fragilisent considérablement le discours utilitariste classique. Dans les chapitres précédents, nous avons évoqué plusieurs types de réponses apportées par les utilitaristes pour en amoindrir les effets, notamment l'excessive exigence morale. Une première forme de réponse consiste à abandonner l'objectif de la maximisation de l'utilité nette, en adoptant par exemple une version modérée de l'utilitarisme (*satisficing*). Cela permet ainsi de justifier les comportements des ménages qui donnent 'un peu' aux organismes caritatifs mais gardent l'essentiel de leur revenu pour l'entretien et l'éducation de leur famille. Une seconde forme de réponse a été apportée par l'utilitarisme de la règle, ou plus exactement par le caractère indirect de l'optimisation de l'utilité qu'il permet. La thèse sous-jacente est qu'il est plus efficace que chaque parent s'occupe de ses enfants et que les amis prennent soin d'eux entre eux, plutôt que nous soyons tous individuellement et impartialement responsables du bien-être des autres<sup>156</sup>. Cette théorie argue qu'elle ne viole pas vraiment le principe d'impartialité puisque cette forme de partialité est généralisable et, de fait, généralisée. Elle reste, cependant, pour certains auteurs, problématique dans la mesure où les relations communautaires sont souvent génératrices de préjugés (religieux, nationaux, sociaux,

---

<sup>156</sup> R.M Hare, *Moral Thinking: Its Levels, Method, and Point*, Oxford: Oxford University Press, 1981.

sexistes, raciaux etc.), qui peuvent être ensuite utilisés pour exclure les minorités de l'aide dont ils auraient pourtant prioritairement besoin<sup>157</sup>.

Le troisième type de réponse est beaucoup plus radical puisqu'il consiste à (partiellement) abandonner le principe de l'impartialité et à lui préférer une approche personnelle (*agent-centered*)<sup>158</sup>. Dans *The Rejection of Consequentialism*, Samuel Scheffler, l'un des principaux partisans d'un conséquentialisme modéré, prend acte du peu d'importance que l'utilitarisme accorde aux personnes, et à leurs désirs<sup>159</sup>. Il défend une approche de la morale recentrée sur la personne (*morality of personal concerns*). Scheffler compare deux alternatives possibles au conséquentialisme. La première repose sur les conceptions totalement centrées sur l'agence (*fully agent-centered*) et désigne implicitement la théorie déontologique. La seconde est une conception hybride qui combine conséquentialisme et reconnaissance morale du statut de l'agent. En intégrant la prérogative de l'agent (*agent-centered prerogative*), la théorie hybride permet de prendre en compte l'indépendance du point de vue de la personne.

Dans les théories conséquentialistes pures, telles que l'utilitarisme, le point de vue de la personne n'est pas complètement ignoré. L'utilité ou la désutilité d'une action pour la personne est prise en compte dans le calcul global d'utilité, mais lorsque celle-ci ne représente qu'une part infime du total, elle est *de facto* négligée. La prérogative de l'agent

---

<sup>157</sup> L. Gruen, "Must Utilitarians Be Impartial?" in Jamieson, D. (ed.), *Singer and his Critics*, Oxford: Blackwell Publishing, 1999, pp.129-149.

<sup>158</sup> T. Mulgan, *Future People. A Moderate Consequentialism Account of our Obligations to Future Generations*, Oxford: Clarendon Press, 2006; S. Scheffler, *The Rejection of Consequentialism: a Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*, Oxford-New York: Clarendon Press-Oxford University Press, 1982.

<sup>159</sup> S. Scheffler, *The Rejection of Consequentialism: a Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*, Oxford-New York: Clarendon Press-Oxford University Press, 1982.

octroyée par la théorie hybride, au contraire, donne un poids *moral* à l'agent, et ce quelle que soit la part de son utilité dans l'utilité totale. Elle reconnaît l'indépendance morale de celui, et l'importance de ses projets ou engagements personnels. Elle permet donc à l'agent d'échapper au calcul impersonnel du maximum d'utilité. Scheffler appelle cela la stratégie de libération (*the liberation strategy*). Elle est opposée à la stratégie de la maximisation (*the maximization strategy*), qui exige que chaque agent agisse toujours de manière à maximiser le nombre de personnes qui réussissent leurs projets de vie<sup>160</sup>. La stratégie de la maximisation reconnaît l'importance de l'indépendance du point de vue personnel dans la mesure où celle-ci contribue au bien individuel, mais elle ne lui donne pas une préséance morale. Scheffler entend ainsi démontrer, contre le déontologisme, qu'il est possible de justifier le sacrifice de l'intérêt d'une personne au profit d'une autre, sans remettre en cause l'intégrité des personnes<sup>161</sup>.

Tim Mulgan s'inscrit dans cette perspective générale mais l'applique plus spécifiquement aux questions intergénérationnelles<sup>162</sup>. Face aux enjeux posés par la surpopulation dans le futur, il considère ainsi que seule une approche centrée sur la personne permet de justifier le fait de pouvoir se reproduire aujourd'hui. L'approche personnelle du conséquentialisme ne justifie pas, cependant, l'égoïsme. Elle s'accompagne également de d'obligations qui nous obligent à justifier nos actions aux personnes qui en seront victimes. Selon Mulgan,

---

<sup>160</sup> S. Scheffler, *The Rejection of Consequentialism: a Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*, Oxford-New York: Clarendon Press-Oxford University Press, 1982, pp.59-60.

<sup>161</sup> J.-F. Spitz, "Le conséquentialisme hybride de Samuel Scheffler : défense et illustration", in S. Guerard De Latour et M.-A. Dilhac (eds.), *Étant donné le pluralisme*, Paris : Éditions de la Sorbonne, 2013, pp.27-53.

<sup>162</sup> T. Mulgan, *Future People. A Moderate Consequentialism Account of our Obligations to Future Generations*, Oxford: Clarendon Press, 2006.

cela donne à nos obligations envers les générations futures une dimension d'urgence que le raisonnement impersonnel est incapable de transmettre<sup>163</sup>.

## B) L'UTILITARISME ET LA REGLE

A l'origine la philosophie utilitariste, telle que conçue par Jeremy Bentham, la décision morale reposait quasi exclusivement sur l'arbitrage rationnel des individus, arbitrage lui-même fondé sur un processus de maximisation. Seule la loi, pensée comme levier à la disposition du législateur utilitariste, introduisait un élément de rigidité dans le processus de décision individuel. A l'inverse, les règles déontologiques, à l'instar de l'impératif catégorique kantien, ne laissaient quant à elles que peu de marge pour l'arbitrage. La radicalité de ces deux systèmes a souvent été soulignée par ses adversaires respectifs, et il est clair qu'un usage quasi exclusif de la règle ou de l'arbitrage se révèle contre-productif en matière de choix moral comme en matière de choix rationnel. L'évolution de la théorie utilitariste classique, aujourd'hui désignée comme utilitarisme de l'acte, vers un utilitarisme de la règle n'a, cependant, pas produit les résultats escomptés. Si elle a permis de justifier l'existence de règles d'un point de vue utilitariste, elle a échoué à montrer à quelles conditions, et selon quelle procédure, on pouvait légitimement s'en exempter. Or, l'obéissance stricte à une règle, quelles que soient les circonstances et les effets attendus de l'action, conduit nécessairement à la réalisation de choix (rationnels ou moraux) que l'on sait être catastrophiques. Le respect absolu de la règle pose également d'insurmontables difficultés en cas de conflits entre règles. Quelle règle suivre, tout d'abord, lorsqu'une même situation peut impliquer différentes règles morales ? Quelle

---

<sup>163</sup> T. Mulgan, "Answering to future people", *Journal of Applied Philosophy*, 35 (3), 2018, p.546.



règle suivre, ensuite, lorsque dans une même situation, nos règles morale et rationnelle nous commandent d'agir différemment ?

L'idée d'un utilitarisme de la règle est développée, et défendue, pour la première fois par Roy Harrod dans son article 'Utilitarianism Revised' (1936). Mais celui-ci eu peu d'influence, et il faudra attendre la seconde moitié du XX<sup>e</sup> siècle pour que la question de la généralisation de l'utilitarisme devienne centrale, notamment avec les travaux de S.E. Toulmin, (*An Examination of the Place of Reason in Ethics*, 1950), M.G. Singer (*Generalization in Ethics*, 1961), David Lyons (*Forms and Limits of Utilitarianism*, 1965) et Richard Brandt (*Ethical Theory : The Problems of Normative and Critical Ethics*, 1959).

La distinction terminologique entre un utilitarisme de l'acte et un utilitarisme de la règle s'est officialisée à partir des années 1950 notamment sous l'impulsion de Richard Brandt en 1959. L'absence de distinction formelle entre ces deux types de conséquentialisme avant le vingtième siècle a conduit de nombreux commentateurs à s'interroger sur la position des fondateurs de l'utilitarisme envers l'un et l'autre. John Stuart Mill est ainsi soit présenté comme un tenant de l'utilitarisme de la règle<sup>164</sup>, soit comme un tenant de l'utilitarisme de l'acte<sup>165</sup> soit comme ne défendant ni l'un ni l'autre<sup>166</sup>. De la même façon, et bien qu'il soit présenté de façon prédominante comme un utilitariste de l'acte<sup>167</sup>

---

<sup>164</sup> Urmson, J. O., "The Interpretation of the Moral Philosophy of J. S. Mill", *The Philosophical Quarterly*, vol. 3, n°10, 1953, pp.33-39.

<sup>165</sup> Crisp, R., *Mill on Utilitarianism*, London: Routledge, 1997; Brink, D.O., *Mill's Progressive Principles*, Oxford: Clarendon Press, 2013.

<sup>166</sup> D. Lyons, *Rights, Welfare, and Mill's Moral Theory*, New York: Oxford University Press, 1994.

<sup>167</sup> De Lazari-Radek, K. et Singer P., *The Point of View of the Universe: Sidgwick & Contemporary Ethics*, Oxford: Oxford University Press, 2014.

Sidgwick est parfois associé à l'utilitarisme indirect<sup>168</sup>. Bentham est, à quelques exceptions près, généralement considéré l'archétype de l'utilitarisme de l'acte.

L'utilitarisme de la règle, dans son acception générale, peut se définir de la façon suivante : « une action est bonne (*right*) si elle est conforme à une règle, qui, *si elle est suivie*, produit au moins autant d'utilité totale que n'importe quelle autre règle possible ». Il se distingue de l'utilitarisme de l'acte que l'on peut définir ainsi « une action est bonne (*right*) si elle produit au moins autant d'utilité totale que n'importe quelle autre action possible ». Dans un cas, le jugement moral sur l'action est déterminé par l'évaluation des conséquences de l'action en question sur le bien-être collectif (*utilitarisme de l'acte*). Dans l'autre, le jugement moral sur l'action est déterminé par l'évaluation des conséquences sur le bien-être général qui seraient produites par la règle à laquelle l'action en question appartient, si cette règle était *normalement* suivie (*utilitarisme de la règle*). Ainsi, si un mensonge, par exemple, peut se justifier par l'utilité nette qu'il peut produire (*utilitarisme de l'acte*), il ne peut pas se justifier moralement s'il est prouvé que la règle 'dire la vérité' produit au moins autant d'utilité nette que la règle 'ne pas dire la vérité'.

Il est important de distinguer plusieurs conceptions de la règle. Une règle peut, tout d'abord, désigner une *pratique* morale et/ou sociale. Ce type de règle n'a pas de fondement utilitariste explicite mais peut néanmoins contribuer *de facto* à maximiser l'utilité. Une règle peut, également, désigner une *heuristique*, c'est-à-dire un raccourci cognitif qui permet d'éviter un calcul de type conséquentialiste. Il ne s'agit donc pas nécessairement

---

<sup>168</sup> B. Schultz, *Henry Sidgwick - Eye of the Universe: An Intellectual Biography*, Cambridge: Cambridge University Press, 2004.

de règles morales. Ces règles, aussi appelées 'règles de pouce' (*rules of thumb*), jouent un rôle prépondérant dans la prise de décision. Leur fonctionnement a beaucoup été étudié, notamment grâce aux travaux d'Herbert Simon et plus récemment d'Amos Tversky et de Daniel Kahneman. Dans la mesure où leur efficacité est attestée (même si elles ne sont pas nécessairement optimales), et qu'elles permettent aux individus d'éviter des stratégies plus coûteuses (en temps, en énergie, en information) et plus risquées (anticipations erronées etc), elles se justifient du point de vue de l'utilitarisme de l'acte. Enfin, une règle peut désigner un *principe* normatif de portée universelle (ex. 'tu ne tueras point') ou de portée restreinte (ex. 'à moins que tu sois en situation de légitime défense, tu ne tueras point'). C'est généralement à cette troisième acception que l'utilitarisme de la règle, au moins dans ses formes primitives<sup>169</sup>, fait référence. Les formes plus contemporaines font plutôt référence aux pratiques<sup>170</sup> ou à l'articulation entre règles théoriques et pratiques<sup>171</sup>.

Les premières formes d'utilitarismes de la règle reposent sur une simple généralisation du raisonnement utilitariste classique: « Que se passerait-il si tout le monde faisait pareil ? ». Le questionnement est légitime puisqu'on a pu reprocher à l'utilitarisme de l'acte de justifier des actions, qui, si elles étaient généralisées, auraient des effets catastrophiques. C'est cette problématique qui a conduit Harrod en 1936 à proposer une nouvelle version de l'utilitarisme. Selon Harrod, le calcul maximisateur, pour être correct, devait penser l'utilité de l'action, non pas simplement en tant qu'utilité d'un acte isolé, mais en tant qu'utilité d'une instance d'un type d'action. L'utilité d'une action n'est, en effet, pas

---

<sup>169</sup> R. F. Harrod, "Utilitarianism revised", *Mind*, vol.45, n°178, 1936, pp.137-156.

<sup>170</sup> B. Hooker, *Ideal Code, Real World: A Rule-consequentialist Theory of Morality*, Oxford : Oxford University Press, 2000.

<sup>171</sup> R.M. Hare, *Moral Thinking: Its Levels, Method, and Point*, Oxford: Oxford University Press, 1981.

indépendante de l'utilité générale du type d'action auquel elle appartient. Autrement dit, une promesse n'a de valeur pour ceux qui la font/reçoivent que parce que la promesse, en tant que type d'action, a une forte utilité sociale. Inversement, les conséquences du non-respect d'une promesse ne peuvent être mesurées que vis-à-vis des conséquences de cette seule action mais doit également intégrer l'impact que cette violation peut avoir sur l'utilité sociale de la promesse en tant que type d'action.

Harrod propose d'abandonner la version 'grossière' (*crude*) de l'utilitarisme et d'y intégrer la formule kantienne de l'universalisation. Il définit le principe de sa version 'raffinée' (*refined*) de l'utilitarisme ainsi : « Il existe certaines actions qui, lorsqu'elles sont réalisées en  $n$  occasions similaires ont des conséquences qui sont plus que  $n$  fois supérieures à celles produites lors d'une seule de ses occurrences »<sup>172</sup>. Une action particulière qui a une utilité nette positive mais qui, répétée  $n$  fois dans des circonstances similaires, aurait une utilité nette totale négative doit donc être considérée comme moralement 'mauvaise' (*wrong*). Cette thèse n'implique pas nécessairement, cependant, que les règles morales doivent être systématiquement obéies. Tout d'abord, il est théoriquement possible que l'infraction répétée d'une règle ait une utilité totale nette. Ensuite, et surtout, ce principe permet de distinguer un certain nombre de circonstances pour lesquelles il est bon d'enfreindre la règle. On peut ainsi à la fois maintenir qu'il est bon de dire la vérité et en même temps considérer qu'il est possible (voire souhaitable) de mentir aux enfants ou aux personnes mourantes.

Cette conception de l'utilitarisme, également adoptée par Marcus Singer, a été très critiquée. Dès 1916, Charlie Dunbar Broad (plus connu sous le nom de C.D. Broad), conteste

---

<sup>172</sup> R. F. Harrod, "Utilitarianism revised", *Mind*, vol.45, n°178, 1936, p.48.

le raisonnement selon lequel on devrait intégrer dans le calcul de maximisation les conséquences d'actions collectives irréelles (et dont la plausibilité n'est pas même avérée). Broad appelle ceci le principe de fausse universalisation et dénonce l'usage des fausses hypothèses en éthique<sup>173</sup>. Selon Broad, l'esprit de la philosophie utilitariste est fondamentalement empirique et scientifique et il est illusoire de fonder ses calculs sur des scénarios fictionnels voire surréalistes. Il est alors absurde de se poser la question « que se passera-t-il si tout le monde fait la même chose ? » sachant que ce ne sera jamais le cas.

La seconde objection portée à la généralisation de l'utilitarisme a été portée par David Lyons en 1965 dans son ouvrage *Forms and Limits of Utilitarianism*. Selon Lyons, la généralisation de l'utilitarisme est une forme 'primitive' de l'utilitarisme de la règle et elle ne doit pas être confondue avec ce dernier. Elle ne constitue pas une rupture épistémologique avec l'utilitarisme de l'acte puisqu'elle est une simple extension logique. Si la maximisation d'utilité justifie le fait que je mente dans la situation x (par exemple à mon enfant pour Noël ou à mon oncle mourant), alors il existe nécessairement une règle selon laquelle il n'est pas permis de mentir sauf dans les toutes les situations de type x. Selon cette version de l'utilitarisme de la règle, il existe autant de règles que situations ou d'actions types. Derrière chaque règle, se profile en fait la maximisation d'utilité d'une action particulière mais généralisée. En d'autres termes, il n'y a pas de différences structurelles entre une généralisation utilitariste et l'utilitarisme de l'acte : les deux théories sont 'par extension équivalentes' (*extensionally equivalent*).

---

<sup>173</sup> C.D. Broad, "On the Function of False Hypotheses in Ethics," *International Journal of Ethics*, vol.26, n°3, 1916, pp.377–97.

Les formes contemporaines d'utilitarisme de la règle se rapportent plutôt à un ensemble de règles plutôt qu'à une règle en particulier. On parle alors de *code moral*. Ce code moral peut être *effectif*, c'est-à-dire existant et socialement accepté, ou *idéal*, autrement dit optimal au regard de l'utilité totale (*optimific*). Le code idéal est un ensemble de règles idéales qui sont acceptées ou internalisées par la très grande majorité de la population. Richard Brandt et Brad Hooker sont considérés comme les principaux théoriciens de l'utilitarisme du code idéal<sup>174</sup>. Les théories du code *réel* (*actual*), moins répandues, ne prennent en compte que les règles déjà existantes<sup>175</sup>. De ce fait, elles ne s'inscrivent pas dans une démarche maximisatrice mais appartiennent plutôt à l'utilitarisme modéré (*satisficing*).

Enfin, certains auteurs considèrent que le raisonnement conséquentialiste ne devrait pas se limiter à un ensemble de règles ou de pratiques, mais devrait également inclure les dispositions ou les motivations des individus. On parle alors plutôt d'utilitarisme *global*<sup>176</sup>. Ainsi, selon David Brink, « [T]out objet d'une évaluation morale doit être évalué, non pas au regard de la valeur de ses conséquences pour le bonheur général, mais par sa conformité à quelque chose d'autre (i.e. des normes, des motifs ou des réactions) dont la valeur d'adhésion (*acceptance value*) est bonne ou optimale<sup>177</sup>. L'utilitarisme global se veut beaucoup plus réaliste que l'utilitarisme de la règle *stricto sensu*. Il permet également de

---

<sup>174</sup> B. Hooker, *Ideal Code, Real World: A Rule-consequentialist Theory of Morality*, Oxford : Oxford University Press, 2000; R.B. Brandt, *A Theory of the Good and Right*, Oxford: Clarendon Press, 1979.

<sup>175</sup> R. B. Miller, "Actual Rule Utilitarianism", *The Journal of Philosophy*, Vol. 106, No. 1, 2009, pp. 5-28; S.E. Toulmin, *An Examination of the Place of Reason in Ethics*, Cambridge: Cambridge University Press, 1950.

<sup>176</sup> J. Driver, "Global Utilitarianism", in B. Eggleston & D. Miller (Eds.), *The Cambridge Companion to Utilitarianism*, Cambridge: Cambridge University Press, 2014, pp.166-176.

<sup>177</sup> D.O. Brink, "Mill's Ambivalence about Rights", *Boston University Law Review*, vol.90, 2010, pp.1669-1704.

mieux prendre en compte la dimension collective de l'action éthique. Cet aspect est particulièrement important dans les débats sur la justice intergénérationnelle, où chaque individu ne peut prendre la pleine responsabilité morale du réchauffement climatique, par exemple.

L'utilitarisme de la règle tente de réconcilier la doctrine utilitariste avec la morale du sens commun, d'une part, et avec la déontologie kantienne, de l'autre. L'utilitarisme du code moral est, de toutes les versions, celle qui a probablement le mieux réussi cette transformation. Mais, ainsi que nous venons de le voir, celle-ci n'est pas sans provoquer des remous parmi les utilitaristes. Ne serait-on pas aller trop loin du côté de la déontologie ? Aurait-on laissé de côté l'essence rationaliste de l'utilitarisme classique ? C'est le sens de la critique de Smart.

Mais au-delà de cette question très théorique s'en pose une autre, plus pratique, aujourd'hui. L'utilitarisme de la règle a souhaité montrer que l'utilitarisme n'impliquait pas forcément des sacrifices excessifs de la part des individus. Les théories du code moral idéal tolèrent même qu'une partie de la population y déroge totalement. Mais n'est-ce pas tomber là dans l'excès inverse ? Cela ne revient-il pas à déresponsabiliser les individus ? La question est posée par un certain nombre de théoriciens et de militants qui dénoncent l'attitude morale qui consiste à ne pas suffisamment se poser la question de notre responsabilité individuelle et à se défaire sur la responsabilité collective. Le renouvellement de l'éthique appliquée, centrée aujourd'hui sur les questions de bioéthique, de souffrance animale et de réchauffement climatique, a probablement beaucoup contribué à cette réévaluation de la responsabilité morale. Elle ne se traduit pas

(encore) par un regain de l'utilitarisme de l'acte, mais certains développements s'orientent dans cette direction.

Le premier d'entre eux est, peut-être, le pessimisme affiché par certains théoriciens de l'utilitarisme de la règle, dont Tim Mulgan, qui vont jusqu'à admettre que le principe d'obéissance partielle (*partial compliance*) ne permet pas à l'utilitarisme de la règle de se constituer une théorie morale complète<sup>178</sup>. Mulgan propose de diviser la morale entre deux branches, le domaine de la nécessité et celui de la réciprocité<sup>179</sup>. Les situations de famine, par exemple, relèvent du domaine de la nécessité. Les situations de la vie ordinaire relèvent, quant à elles, plutôt du domaine de la réciprocité. Mulgan admet que l'utilitarisme de la règle ne permet pas de répondre efficacement à toutes les questions éthiques, et que, dans certains cas, les individus doivent faire le maximum possible pour aider les autres.

Un deuxième développement théorique récent pourrait être interprété comme un signe d'affaiblissement de l'utilitarisme de la règle. Il s'agit est de la théorie de l'altruisme efficace (*effective altruism*). L'altruisme efficace est un mouvement né dans les années 2000 sous l'impulsion conjointe de philosophes utilitaristes, tels que Peter Singer, William MacAskill ou Toby Ord, et de militants du milieu associatif<sup>180</sup>.

La théorie de l'altruisme efficace est assez récente mais est déjà très en vogue. Ses principaux théoriciens sont le philosophe utilitariste Peter Singer, William MacAskill et Toby

---

<sup>178</sup> T. Mulgan, *Future People. A Moderate Consequentialism Account of our Obligations to Future Generations*, Oxford: Clarendon Press, 2006, p.302.

<sup>179</sup> T. Mulgan, *The Demands of Consequentialism*, Clarendon Press, Oxford, 2001.

<sup>180</sup> W. Macaskill, *Doing Good Better: Effective Altruism and a Radical Way to Make a Difference*, London: Guardian Faber, 2015; P. Singer, *The Most Good You Can Do: How Effective Altruism Is Changing Ideas About Living Ethically*, New Haven and London: Yale University Press, 2015.



Ord. Elle invite les agents à adopter une approche totalement rationnelle de l'altruisme, et non seulement prescrit la maximisation du montant des dons mais également la maximisation de leur efficacité, à travers notamment une évaluation stricte de l'usage des dons fait par les associations caritatives. Contrairement à l'usage des organisations humanitaires qui essaient d'abord de minimiser les coûts administratifs, l'altruisme efficace recommande de maximiser la 'rentabilité' de chaque euro donné<sup>181</sup>, et invite tous les donateurs à vérifier le niveau de rentabilité des organisations avant d'effectuer un don. Ils les invitent également à maximiser leurs dons, prescrivant de donner au moins 10% de leurs revenus aux organisations caritatives. Enfin, ils les encouragent à ne pas eux-mêmes s'engager dans des actions humanitaires s'ils ont la possibilité de gagner de l'argent, qui pourra être mieux utilisé dans les mains d'une organisation efficace. Cela peut également impliquer de toujours préférer les carrières à fortes rémunérations, même si celles-ci sont par ailleurs peu satisfaisantes.

La philosophie utilitariste, on l'a vu, fait peser une responsabilité exagérée sur les épaules de chaque individu, un peu comme s'il appartenait à chaque agent de maximiser le bonheur de tous. Elle donne également aux raisons un poids normatif très important. Le fait qu'il existe de bonnes raisons d'augmenter le bonheur de tous, ou tout du moins à minimiser leurs souffrances, n'implique pas que chaque individu soit dans l'obligation morale de se sacrifier dans ce but. Chaque individu doit, cependant, y assumer sa part. Les modalités de sa participation lui appartiennent et lui sont personnelles. Répondre à une raison objective

---

<sup>181</sup> L'impact des actions caritatives est calculé grâce à la méthode de l'essai randomisé contrôlé et des essais randomisés contrôlés conduits par J-PAL (Abdul Latif Jameel Poverty Action Lab), laboratoire du MIT fondé en 2003 par Abhijit Banerjee, Esther Duflo et Sendhil Mullainathan. Cette méthode a néanmoins fait l'objet d'un certain nombre de critiques, voir notamment Judith Favereau, « The J-PAL's experimental approach in development economics: an epistemological turn? », *Erasmus Journal for Philosophy and Economics*, 2014, 7 (2):177-180.

– par exemple éviter les souffrances humaines ou animales – n’implique pas d’y répondre de façon impersonnelle. Bien au contraire. C’est, de mon point de vue, l’une des erreurs de l’altruisme efficace, dont Peter Singer est l’un des représentants les plus en vue.

Le rapprochement de l’utilitarisme de la règle et du contractualisme pourrait, enfin, être le signe d’une évolution, sinon d’un affaiblissement, de l’utilitarisme de la règle. Le contractualisme est, en effet, plus contraignant que le code moral. Les règles sont pensées et acceptées par l’individu, et ce dernier a obligation de les respecter. La règle d’universalisation kantienne, intégrée par Parfit, a par ailleurs pour conséquence de construire chaque règle à partir de la généralisation d’une action. De ce fait, l’utilitarisme de la règle de Parfit se conforme davantage à une généralisation utilitariste que du code moral idéal. Or, ainsi que Lyons l’a montré, la généralisation utilitariste a une structure logique identique à celle de l’utilitarisme de l’acte. Le retour de ce que Lyons a qualifié de forme ‘primitive’ de l’utilitarisme de la règle peut surprendre. Mais, malgré ses nombreuses limites, la généralisation utilitariste a pour avantage de ne pas diluer la responsabilité morale des agents dans une pratique collective.

Parfit développe l’idée selon laquelle les différentes théories que sont la déontologie kantienne, le contractualisme (scanlonien) et l’utilitarisme de la règle ne sont, en fait, que trois approches différentes d’une même ‘réalité’ morale, à l’image de ce que peuvent être les trois faces d’une même montagne, et que l’on a donc tort de les opposer l’une à l’autre. Notons immédiatement que Parfit n’emploie pas l’expression utilitarisme de la règle à laquelle il préfère celle de conséquentialisme de la règle. Contrairement aux philosophies déontologiques et contractualistes, qu’il associe respectivement à Kant et Scanlon, il est

impossible d'identifier avec certitude les fondements théoriques du conséquentialisme de la règle qu'il semble défendre.

Parfit présente, tout d'abord, une version 'kantienne' du conséquentialisme de la règle qu'il définit ainsi : « Chacun devrait suivre les principes optimifiques, car ce sont les seuls principes que chacun pourrait rationnellement vouloir qu'ils deviennent des lois universelles »<sup>182</sup>. Il montre, ensuite (*Argument de la convergence*), que les principes que personne ne pourrait raisonnablement rejeter sont les mêmes que ceux que tout le monde pourrait rationnellement vouloir<sup>183</sup>, et conclut logiquement que ces trois ensembles théoriques peuvent se reformuler sous une seule thèse, qu'il nomme Triple Théorie, et qui établit qu'une action est mauvaise quand elle appartient à un type d'action n'est pas permis par quelque principe qui est optimifique, universellement souhaitable par chacun (*uniquely universally willable*), et qui ne peut être raisonnablement rejeté par personne<sup>184</sup>.

Ce que Parfit nomme Triple Théorie constitue une formulation de l'utilitarisme de la règle, ou plus exactement de l'utilitarisme de la règle dans sa version primitive (au sens de Lyons), à savoir l'utilitarisme de l'acte généralisé. Cela peut étonner puisque, on l'a vu, cette a été beaucoup critiquée dans le passé, et que la plupart des auteurs utilitaristes récents de Harsanyi à Brandt en passant par Hare ont tenté, chacun à leur façon, de montrer que l'utilitarisme de la règle ne se réduisait pas à un utilitarisme de l'acte. Les raisons qui ont poussé Parfit à revenir à cette version primitive de l'utilitarisme de la règle sont essentiellement d'ordre méta-éthiques<sup>185</sup> [Salvat, 2019b]. Mais ce retournement est peut-

---

<sup>182</sup> D. Parfit, *On What Matters*, Oxford: Oxford University Press, 2011, vol.1, p.413.

<sup>183</sup> D. Parfit, *On What Matters*, *op. cit.*, vol.1, p.412.

<sup>184</sup> D. Parfit, *On What Matters*, *op. cit.*, vol.1, p.413.

<sup>185</sup> C. Salvat, "Parfit et les raisons du conséquentialisme de la règle", *Revue de Métaphysique et de Morale*, n°2, 2019, pp.39-50.

être également la marque d'une évolution plus profonde de la pensée utilitariste contemporaine, qui à l'instar de Peter Singer, assume de plus en plus la responsabilité morale *directe* qu'ont les agents lorsqu'ils agissent ou s'abstiennent d'agir.

L'utilitarisme de la règle de Parfit ne peut, par ailleurs, pas être réduit à un utilitarisme de la règle du type généralisation, puisqu'il s'appuie sur un calcul intégrant raisons déontique et raisons non-déontiques. Le conséquentialisme de la règle ne garantit pas, dans sa version traditionnelle, le respect des raisons déontiques. Le conséquentialisme de la règle juge de la valeur de la règle uniquement en fonction de sa contribution à la maximisation du bien total. Il est bon de dire la vérité car on peut raisonnablement penser que dire la vérité est, en général, une règle optimifque. Supposons un instant, cependant, que nous vivions dans une société qui valorise les 'fake news' et que l'on attache plus d'importance à ce que l'on veut croire qu'à ce qui est vrai. On aurait alors tort, dans ce contexte, de dire la vérité. Cette position n'est, cependant, pas conforme à la pensée de Parfit, pour lequel des raisons déontiques peuvent toujours contre-balancer des raisons non-déontiques. Ainsi, explique Parfit, « on peut raisonnablement penser que, dans de nombreux cas, il serait mal en soi de traiter les gens d'une certaine façon, telle que de les tromper ou de les contraindre, ou en violant les promesses qu'on leur a faites. Le mal intrinsèque de telles actions surpasserait souvent le bien de leurs conséquences»<sup>186</sup>. Il ne peut donc exister pour Parfit de règles optimifques fondamentalement contraires aux raisons déontiques (même si des conflits occasionnels sont possibles). Or, rien, en théorie, n'interdit au conséquentialisme de la règle de promouvoir une règle 'immorale' si celle-ci est optimisatrice.

---

<sup>186</sup> D. Parfit, *On What Matters*, Oxford: Oxford University Press, vol.3, 2017, p.421.

Pour répondre à cette objection, Parfit propose donc de ne pas circonscrire le conséquentialisme à l'évaluation des effets de la règle mais de l'appliquer également à ceux des motifs qui lui sont sous-jacents. C'est ce qu'il appelle le 'conséquentialisme de la règle et de la motivation' (*Rule and Motive Consequentialism*)<sup>187</sup>. Si l'impartialité peut se justifier par les effets de la règle (dans la mesure où une égalité de traitement maximise l'utilité totale), la règle selon laquelle il est bon de préférer ses proches aux étrangers se justifie, elle, par l'utilité sociale de la motivation qui l'anime. Un monde dans lequel on n'attacherait pas plus d'importance à ses enfants ou ses parents qu'à de parfaits inconnus ne serait pas un monde heureux. L'optimificité de la règle repose donc conjointement sur les effets de la généralisation de l'action et sur celle de sa motivation. Les uns tempèrent (ou confortent) les autres. A utilité sociale égale, la bonté des motivations d'une règle renforce la raison de s'y soumettre alors que la malveillance inhérente (*intrinsic badness*) l'affaiblit.

Le conséquentialisme de la règle et de la motivation présente un autre avantage par rapport au conséquentialisme de l'acte : il permet de mieux appréhender l'imperceptibilité des effets de certaines actions dont l'action répétée est responsable d'immenses dommages. C'est le cas notamment des actions visant à soulager la pauvreté dans le monde ou des actions visant à limiter la pollution et préserver la biodiversité. Les effets de la règle, ajoutés à ceux de la motivation, donnent aux individus de fortes raisons d'agir en ce sens. Mais ces actions, dont la raison est, je le rappelle, donnée par l'objet (ici le soulagement de la pauvreté et la sauvegarde de la planète) ont-elles véritablement besoin d'être justifiées par les conséquences de leur généralisation ? Cela ne me semble pas évident. L'action qui consiste à aider son prochain est bonne qu'elle soit suivie par tous ou non. En s'écartant de

---

<sup>187</sup> D. Parfit, *On What Matters, op.cit.* vol.3, p.418.

son cadre théorique de départ, Parfit remet en cause l'objectivité même de sa théorie. Le fondamentalisme des raisons répond déjà bien au problème de l'imperceptibilité des effets des actions individuelles. La question des bonnes actions motivées par un mauvais sentiment (ou inversement) ne se pose pas puisque la raison de l'action est donnée par son objet. La motivation personnelle de l'agent ne rentre pas en ligne de compte pas plus que son caractère moral.

Une grande partie de mes recherches à venir devrait avoir pour objet d'exploiter sa théorie des raisons déontiques et non-déontiques pour analyser le mécanisme d'arbitrage entre différents types de raisons, mécanisme qui – selon nous – ne saurait se réduire à un conséquentialisme de la règle, ou tout du moins, pas dans une simple généralisation de l'utilitarisme. L'enjeu de cette recherche est double, voire triple. Il s'agit, tout d'abord, sur un plan normatif d'identifier les principes afférents à une décision moralement juste dans une situation donnée. Il s'agit, ensuite, de l'appliquer à un certain nombre de cas pratiques existants, et notamment à la délicate question du comportement à adopter devant le réchauffement climatique. D'un point de vue méta-éthique, enfin, ces travaux pourraient me permettre de défendre la pertinence du réalisme moral, dont le discrédit constitue (avec le réalisme scientifique) l'un des enjeux principaux de notre siècle. Cette conviction est un élément clé de ma réflexion et de mon évolution disciplinaire.

## 2. Méta-éthique et réalisme moral

A court et moyen terme, je souhaite élargir mes travaux de recherche à l'éthique appliquée et à la méta-éthique, dont j'estime l'articulation avec l'éthique normative fondamentale.

Les questions méta-éthiques, et notamment celle du réalisme moral, font tout d'abord partie de mes travaux sur la philosophie de Parfit, qui y a consacré l'essentiel de *On What Matters*. Je souhaite néanmoins me consacrer prioritairement à l'articulation du niveau méta-éthique et du niveau éthique. Plus exactement, je souhaite pouvoir montrer qu'un questionnement éthique, quel qu'il soit, ne peut entièrement faire l'impasse sur une réflexion méta-éthique. L'exemple de l'expérimentation menée par un groupe de psychologues du MIT sur les intuitions morales initiées par des scénarios fictifs mettant en scène des voitures autonomes, exemple cité plus haut, en est un bon exemple. C'est l'une des raisons pour lesquelles je considère qu'il est préférable de mener une réflexion éthique au sein de la discipline philosophique plutôt qu'en dehors. Néanmoins, même au sein d'une même discipline, l'articulation entre les différents niveaux de l'éthique peut être problématique.

Un premier exemple nous est donné par *On What Matters* de Parfit dans lequel celui-ci est amené, me semble-t-il, à défendre une version obsolète de l'utilitarisme de la règle simplement par celle-ci, étant une simple généralisation de l'utilitarisme de l'acte, est compatible avec (sa lecture de) Kant. Le réalisme moral, adopté par Parfit, l'enjoint en effet soit à prouver que les trois théories éthiques que sont la déontologie kantienne, le contractualisme scanlonien et le conséquentialisme de la règle sont trois formulations différentes d'une même philosophie, soit au contraire à démontrer que seule l'une d'entre elle est vraie. En choisissant, pour se conformer à sa posture méta-éthique, la première

solution, Parfit n'a eu d'autres choix que d'adopter ce que David Lyons comme la version la plus grossière de l'utilitarisme de la règle, au risque de décrédibiliser tout son édifice théorique. S'il est préférable de tenir des positions éthiques et méta-éthiques qui soient compatibles, il peut parfois être regrettable de voir son raisonnement fragilisé simplement pour assurer une bonne articulation entre ces deux niveaux théoriques.

a) CADRE META-ETHIQUE : LE FONDAMENTALISME DES RAISONS

Les raisons, nous dit Parfit, justifient les normes. Il existe traditionnellement deux grandes approches selon que l'on considère ou non le sujet comme la justification ultime de la norme. Dans le premier cas, nos raisons sont relatives aux faits satisfaisants (ou appelés à satisfaire) nos désirs. Dans la mesure où l'importance de ces faits n'est relative qu'à nos propres désirs, et donc à notre propre personne, on peut selon Parfit également les appeler raisons données par-le-sujet (*subject-given reasons*). Leurs forces dépendent de la force de nos désirs. Dans le second cas, les raisons sont relatives aux faits qui concernent les objets de nos désirs. On parle alors de raisons données-par-l'objet (*object-given reasons*). Leurs forces dépendent des faits qui nous donnent des raisons de les désirer. Ajoutons à ces deux catégories un troisième type de raisons données-par-l'état (*state-given reasons*), qui à la différence des raisons données-par-l'objet ne sont pas relatives aux faits qui concernent les objets de nos désirs mais aux états constitués par le fait d'avoir ces désirs<sup>188</sup>. Parfit, à la

---

<sup>188</sup> D. Parfit, *On What Matters*, Oxford: Oxford University Press, 2011, vol.1, p.50.



différence de Scanlon notamment<sup>189</sup>, considère cette dernière catégorie comme secondaire<sup>190</sup>.

Quelques précisions sont nécessaires ici. Affirmer qu'une théorie des raisons est objective ne signifie pas, tout d'abord, qu'il s'agisse d'une théorie des raisons objectives. Dans le premier cas, l'objectivité des raisons renvoie à un jugement extérieur porté sur les raisons de l'individu alors que, dans le second, l'objectivité renvoie à l'extériorité des raisons par rapport au sujet. Parfit confond parfois ces deux perspectives. A strictement parlé, les raisons-données-par-l'objet ne sont pas nécessairement objectives dans le sens où elles font l'objet d'une évaluation totalement décontextualisée. Les raisons-données-par-l'objet ne sont pas *subjectives*, au sens où elles sont factuelles, c'est-à-dire qu'elles ont trait aux propriétés de l'objet du désir plutôt qu'au désir lui-même, mais elles n'en sont pas moins *personnelles*, c'est-à-dire que leur valeur dépend du rapport entre l'objet et la personne ainsi que les circonstances dans lesquelles l'évaluation de la force de la raison a lieu. Considérons le fait de manger une pomme, par exemple. Selon la théorie des raisons-données-par-le-sujet, j'ai une raison de manger une pomme si je désire manger une pomme. Les caractéristiques de l'objet (ici la pomme) n'entrent pas en ligne de compte dans la justification de mon action 'manger la pomme'. Je pourrais remplacer ici 'pomme' par 'terre' et justifier mon action de manger de la terre par le fait que je désire manger de la terre. Parfit rejette cette approche. Il considère, au contraire, que ce sont les qualités intrinsèques de la pomme, autrement dit ses qualités nutritives et gustatives qui justifient

---

<sup>189</sup> Scanlon considère, en effet, que l'évaluation des biens substantiels (il rejette la terminologie de bien objectif) repose, en grande partie, sur la valeur des états de conscience désirables (desirable states of consciousness). *Value, desire and Quality of Life*.

<sup>190</sup> D. Parfit, *On What Matters, op.cit.*, vol.1., p.51.

l'action 'manger une pomme'. Mais, bien évidemment, cette raison ne s'applique pas également à tous et dans toutes circonstances. Certaines personnes n'apprécient pas ses qualités gustatives, n'ont tout simplement pas faim (auquel cas la qualité nutritive de la pomme est sans objet) ou encore que la pomme, dans certaines circonstances, n'est pas comestible ou digeste. Dans ces cas-là, comme dans d'autres, il n'est pas justifié pour la personne en question de manger cette pomme (même s'ils désirent la manger).

Ce point est, il me semble, important car il permet de mieux comprendre la nature exacte de l'hédonisme parfitien. Parfit, rappelons-le, distingue trois grandes approches de la valeur : l'hédonisme (selon laquelle ce qui est bon est ce qui donne du plaisir aux gens), la théorie de la satisfaction des désirs (selon laquelle ce qui est bon est ce qui satisfait les désirs d'une personne tout au long de sa vie), et la théorie de la liste objective (selon laquelle certaines choses sont bonnes que nous les désirions ou non)<sup>191</sup>. Dans OWM, Parfit – on l'a dit – rejette explicitement toute approche subjective des raisons<sup>192</sup>. Il rejette donc ce qu'il appelle les raisons-relatives-aux-désirs (*desire-based reasons*) au profit des raisons relatives-aux-valeurs (*value-based reasons*). Mais, et c'est un point essentiel, le rejet des raisons relatives au désir n'implique pas le rejet de l'hédonisme dans la mesure où les raisons hédonistes ne sont pas relatives aux désirs que l'on a mais aux états de plaisirs produits par l'objet du désir<sup>193</sup>. Elles appartiennent donc bien aux raisons-données-par-l'objet. L'argument dit de l'agonie (*the Agony Argument*) pourrait faire croire que Parfit embrasse une théorie de la valeur objective. Ce n'est pas le cas. L'argument de l'agonie, selon lequel nous avons tous des raisons de vouloir éviter une agonie future, vise à

---

<sup>191</sup> D. Parfit, *Reasons and Persons*, *op. cit.*, p.4.

<sup>192</sup> D. Parfit, *On What Matters*, *op.cit.*, vol.1., p.65.

<sup>193</sup> D. Parfit, *On What Matters*, *op.cit.*, vol.1., p.67.

décrédibiliser les raisons-relatives-aux-désirs (ce que Parfit appelle également le *subjectivisme*), et non pas l'hédonisme. Il est ainsi tout à fait possible d'associer des raisons relatives-aux-valeurs (ce que Parfit appelle *l'objectivisme*) avec des raisons hédonistes données-par-l'objet. Dans une telle perspective, même si on ne se soucie pas de notre futur bien-être, on a des raisons de s'en soucier<sup>194</sup>. Le fait que la souffrance soit mauvaise en tant qu'elle implique des raisons de l'éviter ne signifie pas que la souffrance soit objectivement mauvaise, c'est-à-dire qu'elle puisse être jugée comme telle par une autorité extérieure<sup>195</sup>. La dimension objective de l'hédonisme de Parfit fait donc ici uniquement référence à la qualité des objets du désir de l'individu et non à l'extériorité du jugement. C'est la raison pour laquelle son rejet du subjectivisme n'implique pas son adhésion à la théorie de la liste objective.

Ceci nous permet également de mieux comprendre la distinction établie par Parfit entre le caractère *impersonnel* et le caractère *impartial* des raisons. L'impartialité, en philosophie, est une qualité du jugement moral généralement associée à un point de vue objectif, c'est-à-dire extérieur à l'individu, ce que l'on appelle le 'point de vue de l'univers' (Sidgwick) ou le 'point de vue de nulle part' (Nagel). Du fait de cette association, on oppose généralement ce qui est personnel à ce qui est impartial. Pour Parfit, cependant, l'objectivité renvoie, ainsi qu'on l'a dit, à l'objet du désir et non au point de vue adopté. Le caractère personnel d'une raison, c'est-à-dire le fait qu'elle est considérée comme raison d'agir par une personne, n'est par conséquent pas un élément déterminant pour juger de son impartialité, c'est-à-dire de sa conformité à la morale. Il est donc bon de distinguer les

---

<sup>194</sup> D. Parfit, *On What Matters, op.cit.*, vol.1., p.74.

<sup>195</sup> D. Parfit, *On What Matters, op.cit.*, vol.1., pp.371-72.

raisons *personnelles* (c'est-à-dire les raisons qui s'imposent aux personnes) des raisons *relatives aux personnes* (c'est-à-dire des raisons ayant spécifiquement trait au bien-être de la personne en question ou à celui de personnes qui lui sont chères)<sup>196</sup>. Pour le dire plus simplement, les raisons personnelles ne sont pas nécessairement intéressées alors que les raisons partiales le sont.

L'idée selon laquelle les raisons peuvent à la fois être personnelles et impartiales est centrale dans l'objectif qui est le sien de résoudre le dilemme du dualisme de la raison pratique de Sidgwick. Ce dernier, rappelons-le, considérait que nous avons à la fois des raisons d'agir pour le mieux, pensé de façon impartiale, et des raisons des raisons d'agir pour ce que nous considérons être notre propre bien, mais que les unes et les autres ne sont pas comparables. Nous avons, en effet, des raisons suffisantes d'agir dans un sens comme dans l'autre, mais aucune raison décisive d'agir dans un sens plutôt que dans un autre. Les deux alternatives, qui expriment respectivement le point de vue objectif (ou, selon l'expression de Sidgwick, 'le point de vue de l'Univers') et le point de vue subjectif, sont toutes deux parfaitement rationnelles.

Pour Parfit, au contraire, les deux points de vue sont concomitants. Le fait que j'adopte un point de vue personnel n'exclut pas le fait que je puisse penser de façon impartiale. Adopter un *point de vue impartial* n'implique pas que *l'on* soit impartial<sup>197</sup>. Le point de vue personnel n'est en effet pas réductible à l'égoïsme rationnel. Je suis *à la fois* capable de désirer être

---

<sup>196</sup> D. Parfit, *On What Matters, op.cit.*, vol.1., p.40.

<sup>197</sup> D. Parfit, *On What Matters, op.cit.*, vol.1., p.41. Inversement, la subjectivité n'est pas incompatible, pour Parfit, à l'impartialité des raisons. Je peux reconnaître, en tant que sujet, que ma vie ne vaut pas mieux que celle qu'un autre, D. Parfit, *On What Matters, op.cit.*, vol.1., p.139.

sauvé et de désirer que le maximum de personnes soit sauvé. Le fait que ces raisons diffèrent, voire rentrent en conflit, ne justifie pas de conclure, à l'instar de Sidgwick, à l'indétermination du choix. La principale difficulté, pour Parfit, consiste cependant à établir les termes de la comparaison entre les raisons qui sont causées par l'objet d'un désir personnel et celles qui ne le sont pas, notamment les raisons que nous imposent notre code moral.

La théorie des raisons adoptée par Parfit doit, on l'a déjà souligné, beaucoup à Scanlon. Or, le philosophe de Harvard conteste le fait que le bien-être individuel puisse être présenté comme le facteur unique, voire même primordial, auquel se réduirait *in fine* toutes les raisons génériques. Une telle approche des raisons conduirait à développer ce qu'il nomme un contractualisme du bien-être (*welfarist contractualism*) auquel il refuse explicitement d'être associé. Parmi les raisons que l'hédonisme ne permet pas d'appréhender sont naturellement « les raisons ayant clairement un caractère moral »<sup>198</sup> dont celles ayant trait à la responsabilité individuelle et à la justice (*fairness*). Parfit institutionnalise en quelque sorte cette distinction en distinguant des raisons déontiques, pour lesquelles relèvent les raisons à caractère moral, et les raisons non-déontiques auxquelles appartiennent les raisons qui se justifient par l'hédonisme. La volonté de Parfit de catégoriser, qui en pratique prendra souvent les traits d'une opposition, ces deux types de raisons alors que Scanlon se refusait explicitement à distinguer (et par conséquent à hiérarchiser) le bien-être des autres valeurs personnelles<sup>199</sup>, n'est pas anodin. La similarité syntaxique de ce dualisme n'est, en effet, pas sans rappeler celui formulé par Sidgwick bien qu'ils diffèrent

---

<sup>198</sup> T.M Scanlon, *What We Owe to Each Other*, Cambridge, Massachusetts and London, England: The Belknap Press of Harvard University Press, 1998, p.243.

<sup>199</sup> T.M Scanlon, *What We Owe to Each Other*, *op.cit.*, p.142.

sémantiquement (l'opposition point de vue subjectif-point de vue objectif n'est pas réductible à l'opposition raisons non-déontiques-raisons déontiques). En solutionnant le second, Parfit peut donc donner l'impression de solutionner le premier.

Avant de revenir sur la comparaison entre raisons déontiques et raisons non-déontiques, il est nécessaire de clarifier la procédure de comparaison en jeu au sein de la catégorie des raisons non-déontiques selon qu'elles sont relatives à la personne (c'est-à-dire qu'elles sont intéressées) ou non. Bien que les raisons soient objectives, dans le sens où elles sont données par l'objet, elles ne s'appliquent pas nécessairement de façon égale à tous. C'est une implication directe de sa conception de l'hédonisme développée dans la section précédente. Il en généralise la portée dans ce qu'il appelle la thèse du particulier-au-général (*The Any-All thesis*) selon laquelle « (S)i quelqu'un a une raison personnelle de vouloir que quelque chose arrive, tout le monde a une raison impartiale, *bien que peut-être de façon plus faible*, de vouloir que cette chose arrive également. »<sup>200</sup>. Le fait que je reconnaisse que les raisons d'autrui ne soient pas moins fondées que les miennes ne m'engage pas à les traiter comme les miennes. Elles n'ont pas pour moi la force qu'elles ont pour eux. Si la souffrance<sup>201</sup> constitue toujours une raison de vouloir l'éviter, elle n'a

---

<sup>200</sup> D. Parfit, *On What Matters, op.cit.*, vol.3, p.321. Je souligne.

<sup>201</sup> Notons que la grande majorité des exemples présentés par Parfit évoque la souffrance ou la mort potentielle. C'est le cas notamment (dans OWM) de Lift, de Tunnel, de Sidetrack et de Bridge (qui sont tous des reformulations du célèbre cas du Trolley), de celui la Nuclear Policy (inspiré des bombardements de Hiroshima et Nagasaki), ou encore de ces exemples fictifs aux noms évocateurs : Child Drowning, Shipwreck, Lifeboat, Fire, Earthquakes, Desperate Plight, Whimsical Despot, Murderous Theft, Gangster, Lesser Evil, Depression etc. Rares sont les exemples qui justifient la force d'une raison non-déontique par l'intensité et/ou la quantité de plaisir qu'elle pourrait procurer. Doit-on en conclure que Parfit défend, au moins implicitement, une forme de conséquentialisme négatif ? On peut raisonnablement le penser. Dans *Reasons and Persons*, Parfit affirme néanmoins que « le soulagement de la souffrance devrait avoir un poids plus important que celui qui lui ai ordinairement donné par les Utilitaristes Positifs. Mais il ne semble pas plausible qu'il doive avoir priorité absolue, ou même un poids beaucoup plus important. » D. Parfit, *Reasons and Persons, op. cit.*, pp.344-45.

pas la même force si elle me concerne, ou concerne mes proches, que si elle concerne des inconnus. Cette considération peut, cependant, être contrebalancée par le nombre des personnes en souffrance, en nous amenant, par exemple, à préférer nous sacrifier pour sauver un nombre plus important de personnes, sans que l'on ne puisse jamais, établir un critère de délibération fixe. Les raisons non déontiques, fondées sur la gravité des souffrances ou l'ampleur des plaisirs, sont toujours comparables, mais elles ne le sont qu'*approximativement*.

L'évaluation de la force des raisons déontiques, et par conséquent de leur comparabilité, pose néanmoins plus de problèmes. Les raisons déontiques peuvent être définies comme des devoirs ou des prohibitions de nature non-conséquentialiste. Il en existe plusieurs types. Les premières relèvent de principes ayant valeur (universelle ou non) qui se justifient par le fait qu'ils pourraient être rationnellement choisis et/ou acceptés par tous, selon la version kantienne (*the Kantian Formula*), ou que personne ne pourrait raisonnablement rejeter, selon la version de Scanlon (*the Scanlon Formula*). Mais il peut également s'agir, non de principes, mais de devoirs ou de prohibitions *prima facie*, selon l'expression de David Ross<sup>202</sup>, autrement dit de devoirs auxquels on doit se conformer à moins qu'on ait de meilleures raisons de faire autrement. Les raisons déontiques chez Parfit sont, en quelque sorte, un combiné de ces trois approches. Elles sont personnelles mais ayant néanmoins une dimension générale<sup>203</sup>, et peuvent entrer en conflit les unes avec les autres.

---

<sup>202</sup> W.D. Ross, *The Right and the Good*, Oxford: Oxford University Press, 1930.

<sup>203</sup> J'utilise ici à dessein le terme de général plutôt que celui d'universel. Si les raisons ont clairement une dimension universelle chez Kant, elles n'en ont pas chez Scanlon qui accepte un certain pluralisme des raisons lié notamment aux différences de culture. La position intermédiaire de Parfit à cet égard ne me semble pas très claire. Dans l'incertitude, je préfère donc parler de dimension générale des raisons.

Parfit croit ainsi que « nous avons des raisons personnelles et impartiales qui ne nous sont pas fournies, ou pas totalement fournies, par les faits relatifs à notre bien-être ou celui des autres. Nous avons des raisons d'essayer de faire différentes choses qui ont de la valeur en elles-mêmes, telles que découvrir la vérité, créer ou préserver la beauté, réaliser plusieurs autres buts bons ou valables. Nous avons aussi des raisons morales à la fois neutres et relatives à la personne, comme les raisons de prendre soin de nos propres enfants, et des raisons de faire en sorte que les avantages et les inconvénients soient plus également distribués entre les différentes personnes »<sup>204</sup>.

Comment comparer une raison non déontique, telle que la recherche du confort personnel, à une raison déontique, telle que la préservation de la nature ? Selon que l'on raisonne en conséquentialiste ou en déontologue, on aura tendance à donner systématiquement plus de poids aux raisons déontiques ou, au contraire, aux raisons non déontiques, rendant ainsi factice le dépassement de l'indétermination sidgwickienne. Une possibilité pourrait, cependant, consister à évaluer le poids des raisons déontiques en fonction de la valeur de la règle qui les porte, autrement dit, à donner une raison non déontique aux raisons déontiques. C'est ce que fait le conséquentialisme de la règle, auquel la théorie de Parfit semble conduire.

---

<sup>204</sup> D. Parfit, *On What Matters, op.cit.*, vol.3, p.321.



b) DE L'IMPACT DE LA META-ETHIQUE SUR L'ETHIQUE NORMATIVE : L'EXEMPLE DE LA TRIPLE THEORIE DE PARFIT

La publication d'*On What Matters*, à partir de 2011, apparaît comme un tournant dans l'œuvre de Parfit. D'un style plus académique, *On What Matters* s'ouvre aux considérations méta-éthiques délaissées dans *Reasons and Persons*. Il y développe également une approche très inspirée du contractualisme de Thomas Scanlon. Néanmoins l'adoption par Parfit d'une théorie des raisons objective ne constitue pas à ses yeux une remise en cause de l'approche conséquentialiste. Il préfère voir ces deux approches comme des méthodes complémentaires.

L'approche méta-éthique de Parfit peut se résumer par la formule 'Cognitivism Normatif Non-Naturaliste et Non-Métaphysique' ou encore 'Rationalisme'<sup>205</sup>. Parfit est, tout d'abord, partisan du cognitivism moral, théorie selon laquelle il est possible d'affirmer qu'une proposition morale est vraie ou fausse. Il s'oppose, en cela, aux non-cognitivistes tels que les prescriptivistes (Hare), les expressivistes (Gibbard) et les quasi-réalistes (Blackburn) pour lesquels les propositions morales ont essentiellement une fonction sémantique et expriment un état mental non-cognitif comme par exemple un désir ou des dispositions émotionnelles. Pour Parfit, il existe donc des propositions morales, dont on peut prouver la véracité. Ces propositions morales, une fois identifiées, ont une valeur normative dans la mesure où elles constituent pour l'agent des *raisons* d'agir. On pourrait donc dire, dans un langage ordinaire, qu'il existe des vérités morales qui, une fois connues, peuvent motiver l'action. La propriété non-métaphysique du cognitivism parfitien précise

---

<sup>205</sup> D. Parfit, *On What Matters, op.cit.*, vol.2., p.113 et 486.

l'état (ou dans ce cas l'absence d'état) ontologique de ces vérités morales. Puisque Parfit rejette toute approche naturaliste, ces vérités ne peuvent donc exister en tant que faits.

Plus fondamentale pour notre propos est sa position anti-naturaliste. Le renouveau de la philosophie morale non-naturaliste doit beaucoup à G.E. Moore et à son *Principia Ethica* (1903) dans lequel il dénonçait le sophisme naturaliste de certaines théories utilitaristes. Parfit s'inscrit pleinement dans cette tradition et réfute l'idée selon laquelle les vérités normatives puissent être définies par leurs propriétés naturelles. Historiquement parlant, le naturalisme est né de la volonté des philosophes du XVII<sup>e</sup> et XVIII<sup>e</sup> siècle de s'opposer aux arguments métaphysiques en moral et d'adopter une approche scientifique et/ou empirique en éthique. C'est le cas de l'utilitarisme de Bentham et Mill. Le naturalisme fait, cependant, face à une difficulté théorique fondamentale dans la mesure où il échoue à rendre compte de la normativité de telle ou telle propriété naturelle. Le naturalisme est donc soit trivial (toutes les propositions morales sont des faits naturels) soit faux (il n'existe pas de proposition morale qui soit réductible à une propriété naturelle). Moore opte pour la première option, Parfit pour la seconde. Ainsi que le souligne Laurent Jaffro, la théorie cognitiviste de Parfit est assez originale car elle n'implique pas une attitude réaliste des valeurs et elle permet de renouveler le débat sur le statut des propriétés normatives des raisons<sup>206</sup>.

A l'instar d'autres réalistes moraux, comme Scanlon, il considère donc qu'il existe des normes morales (bien qu'elles n'aient pas d'existence dans le sens ontologique du terme) et que celles-ci ne soient pas réductibles à des faits naturels. Parfit considère que le

---

<sup>206</sup> L. Jaffro, «A qui faire porter le chapeau des valeurs ? Une discussion du fondamentalisme des raisons», *Revue de Métaphysique et de Morale*, 2019, 2, pp.51-64.

réalisme moral est fragilisé par l'existence de plusieurs philosophies morales rivales et que la convergence de ces différentes théories est possible sous certaines conditions<sup>207</sup>. La thèse de la convergence des différentes théories éthiques permet de reconnaître la validité de chacune sans remettre en question l'existence objective de normes. Le conséquentialisme de la règle a essentiellement pour objet d'intégrer l'équation kantienne dans un modèle contractualiste, que l'on peut qualifier après Scanlon, de contractualisme du bien-être (*welfare contractualism*). Il lui permet également de mettre l'accent sur la dimension hédoniste des raisons que Scanlon tend, selon lui, à minimiser. Reste que le modèle des raisons déontiques et non-déontiques développé par Parfit ne peut être présenté comme un conséquentialisme de la règle dans la mesure où l'évaluation de la force des raisons est fonction des personnes auxquelles elles s'imposent. Il ne peut donc pas y avoir de systématisme de la règle dans le modèle de Parfit.

Deux arguments critiques lui ont été opposés. Le premier est celui du 'dilemme darwinien' (*Evolutionary Debunking Argument*). Bien que d'origine assez ancienne, il a trouvé une nouvelle formulation et une nouvelle fortune grâce aux travaux de la philosophe américaine Sharon Street. L'argument, en soi, est simple. S'il est prouvé, comme cela semble être le cas, que nos normes morales ont évolué au cours du temps en se conformant au principe de sélection naturelle<sup>208</sup> (Sober and Wilson, 1998, Kitcher, 2009, de Waal, 2009), alors on ne peut raisonnablement soutenir, comme le font Parfit et les autres

---

<sup>207</sup> D. Parfit, *On What Matters, op.cit.*, vol.2, pp.549-565.

<sup>208</sup> P. Kitcher, «Ethics and Evolution. How to get here from there» In *Primates and Philosophers: How Morality Evolved*, edited by S. Macedo and J. Ober, Princeton and Oxford, Princeton University Press, 2009, p. 120-139; E. Sober et D. Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge, Harvard University Press, 1998; F. De Waal, *Primates and Philosophers: How Morality Evolved*, Princeton and Oxford, Princeton University Press, 2009.

réalistes, qu'il existe des vérités morales indépendantes des attitudes évaluatives des individus. Le second point abordé est celui de « l'argument du désaccord » (*The Argument from Disagreement*). Celui-ci se formule également assez aisément : le fait qu'il persiste des désaccords irrésolubles en matière morale infirme l'idée d'une objectivité des normes morales.

Comment, tout d'abord, concilier le principe de raisons normatives objectives, indépendantes des attitudes évaluatives des individus, avec la théorie de l'évolution, qui montre – au contraire – que l'évaluation des normes a déterminé leur impact sur la survie et la reproduction des individus ? Parfit, précisons-le tout de suite, ne sous-estime pas l'importance de la sélection naturelle en éthique<sup>209</sup>. Il admet que les normes morales qui ont résisté au passage du temps, telles que l'obligation d'aider son prochain, l'interdiction de tuer ou de mentir, présentent un certain nombre d'avantage sur le plan de l'évolution. Elles ont largement contribué à la survie et à la reproduction des communautés humaines, même si dans certains cas elles ont également nourri la compétition entre communautés. Le fait, cependant, qu'une norme soit perçue comme étant bénéfique à la survie de l'espèce humaine n'est pas incompatible avec le fait qu'elle soit vraie. En fait, nous dit Parfit, elle nous est très probablement bénéfique *parce qu'elle est vraie*<sup>210</sup>. L'argument de Parfit repose sur une hypothèse évolutionniste lourde bien qu'intéressante. Selon lui, si les guépards doivent leur survie à leur vitesse, et les girafes à leur long cou, les hommes doivent la leur à leur capacité à répondre aux raisons, et notamment aux raisons

---

<sup>209</sup> Notons, que dans *Reasons and Persons*, Parfit adopte un point de vue beaucoup plus sévère envers les explications évolutionnistes. Le fait qu'un comportement ou une croyance morale soit le produit de milliers d'années de sélection naturelle n'a pour lui aucune incidence morale D. Parfit, *Reasons and Persons, op. cit.*, p.308.

<sup>210</sup> D. Parfit, *On What Matters, op.cit.*, vol.2., p.512.

épistémiques<sup>211</sup>. Cette faculté cognitive a elle-même évolué au cours du temps, et cela explique, pour Parfit, que si les hommes ont se sont parfois trompés sur ce qu'ils croyaient être des vérités normatives, et que leurs croyances ont souvent été détournées de la vérité du fait des forces évolutionnistes, ils sont aujourd'hui beaucoup moins enclins à faire de telles erreurs<sup>212</sup>. L'altruisme pur, la condamnation des discriminations raciales ou sexuelles, voire la protection des animaux, ne s'expliquent pas – selon Parfit – par la sélection naturelle. Pourtant, ce sont là des vérités normatives. Le fait que certaines normes morales objectives puissent également se justifier d'un point de vue évolutionniste ne doit pas nous faire perdre de vue la réalité de l'existence objective de ces normes indépendamment de toute justification contingente.

Que se passerait-il, ensuite, si l'on pouvait démontrer l'existence de deux vérités normatives dissonantes ? Comment croire à la théorie du réalisme moral lorsqu'on ne parvient pas à se mettre d'accord sur la nature de ces vérités morales ? Ce sont les questions que pose l'argument du désaccord. Celui-ci a été formulé pour la première fois par Mackie, qui l'utilisait essentiellement pour contrer l'intuitionnisme moral<sup>213</sup>. Il a, depuis, été reformulé, notamment par Stephen Stich, qui définit ce qu'il appelle un 'désaccord moral fondamental' comme un désaccord qui ne saurait être dissipé, y compris dans des conditions idéales et y compris lorsqu'il y a accord sur les faits non-moraux. A ceci Parfit répond qu'il suffit de pouvoir montrer que, dans des conditions idéales, les individus auraient *presque* tous des croyances normatives *suffisamment proches*<sup>214</sup>. Or, il est possible

---

<sup>211</sup> D. Parfit, *On What Matters, op.cit.*, vol.2., p.528.

<sup>212</sup> D. Parfit, *On What Matters, op.cit.*, vol.2., p.538.

<sup>213</sup> J.L Mackie, *Ethics: Inventing Right and Wrong*, London, Penguin Books Limited, 1977.

<sup>214</sup> D. Parfit, *On What Matters, op.cit.*, vol.2., p.547.

de montrer qu'en dépit des apparences, les différences croyances morales ne se réduisent pas à des désaccords normatifs profonds.

Sa Triple Théorie, qui démontre la convergence des théories utilitaristes, kantienne et contractualistes en est un bon exemple. Une partie importante du premier volume est consacré à l'étude et l'interprétation de la pensée de Kant. Parfit propose une relecture de la loi universelle de Kant qu'il reformule ainsi « Chacun devrait suivre les principes dont l'acceptation universelle pourrait être rationnellement voulue par tous »<sup>215</sup>. Parfit va plus loin dans sa lecture contractualiste de Kant. Il montre que 'les principes dont l'acceptation universelle pourrait être rationnellement voulue par tous' sont en fait des principes optimisateurs. Cela montre, pense-t-il, qu'on a tort d'opposer la pensée déontologique de Kant et la philosophie conséquentialiste des utilitaristes. Selon Parfit, le contractualisme kantien converge logiquement vers l'utilitarisme de la règle, qui est lui-même compatible avec le contractualisme de Scanlon. Ce conséquentialisme kantien de la règle (*Kantian Rule Consequentialism*) est ainsi défini : « Chacun devrait suivre les principes optimifiques, car ce sont les seuls principes que chacun pourrait rationnellement vouloir devenir des lois universelles »<sup>216</sup>. Combiné au contractualisme de Scanlon, on obtient le principe de la Triple Théorie, à savoir que « Une action est mauvaise (*wrong*)<sup>217</sup> quand ce type d'action n'est

---

<sup>215</sup> D. Parfit, *On What Matters, op.cit.*, vol.1., p.342.

<sup>216</sup> D. Parfit, *On What Matters, op.cit.*, vol.1., p.411. Plus exactement, il affirme que nous avons de 'fortes raisons' d'accepter ce principe, D. Parfit, *On What Matters, op.cit.*, vol.1., p.418. Notez l'emploi ici du 'ought' qui joue d'une certaine ambiguïté entre l'idée de devoir rationnel (avoir suffisamment de raisons pour) et celle de devoir moral (être dans l'obligation morale de).

<sup>217</sup> L'usage du mot *wrong* semble faire du principe de la Triple Théorie une règle morale normative en contradiction avec sa théorie des raisons (voir plus bas et Jaffro 2019). Il l'abandonne d'ailleurs dans ses formulations ultérieures du principe de la triple théorie. Dans le troisième volume il écrit ainsi « qu'on devrait agir selon les principes ou règles (*we ought to act on the principles or rules*) qui sont optimifiques, universellement souhaitables par chacun (*uniquely universally willable*), et qui ne peuvent être raisonnablement rejetés » D. Parfit, *On What Matters, op.cit.*, vol.3, p.434.

pas permis par quelque principe qui est optimifique, universellement souhaitable par chacun (*uniquely universally willable*), et qui ne peut être raisonnablement rejeté »<sup>218</sup>.

De la rencontre de ces trois théories morales, Parfit produit une 'triple théorie' selon laquelle un acte est mauvais si et seulement si de tels actes sont proscrits par tout principe qui a) appartient aux principes optimifiques, b) est l'un des seuls principes dont chacun pourrait rationnellement souhaiter l'universalisation et c) est un principe que personne ne pourrait raisonnablement rejeter<sup>219</sup>. Dans la mesure où un tel principe éthique existe, on peut alors dire, avec Parfit, qu'il existe une vérité morale qui transcende les oppositions théoriques. Pour reprendre l'analogie de Parfit, Kant, les contractualistes et les conséquentialistes « escaladent la même montagne sur des versants différents »<sup>220</sup>. Le principe de la Triple Théorie par lequel s'achève le premier volume d'OWM est présenté comme l'aboutissement ultime de sa réflexion éthique.

L'hypothèse que je formule dans un article de la *Revue de Métaphysique et de Morale*<sup>221</sup> est que la Triple Théorie est censée conforter la position méta-éthique de Parfit, qui se décrit lui-même comme un adepte du cognitivisme normatif non-métaphysique et non-naturaliste<sup>222</sup>. A l'instar d'autres réalistes, comme Scanlon, il considère donc qu'il existe des normes morales (bien qu'elles n'aient pas d'existence dans le sens ontologique du terme) et que celles-ci ne soient pas réductibles à des faits naturels. Parfit considère que le réalisme moral est fragilisé par l'existence de plusieurs philosophies morales rivales et que

---

<sup>218</sup> D. Parfit, *On What Matters, op.cit.*, vol.1., p.413.

<sup>219</sup> D. Parfit, *On What Matters, op.cit.*, vol.1., p.412-13.

<sup>220</sup> D. Parfit, *On What Matters, op.cit.*, vol.1., p.419.

<sup>221</sup> C. Salvat, "Parfit et les raisons du conséquentialisme de la règle", *Revue de Métaphysique et de Morale*, n°2, 2019, pp.37-48.

<sup>222</sup> D. Parfit, *On What Matters, op.cit.*, vol.2., p.486.

la convergence de ces différentes théories est possible sous certaines conditions<sup>223</sup>. La thèse de la convergence des différentes théories éthiques permet de reconnaître la validité de chacune sans remettre en question l'existence objective de normes. Le conséquentialisme de la règle a essentiellement pour objet d'intégrer l'équation kantienne dans un modèle contractualiste, que l'on peut qualifier après Scanlon, de contractualisme du bien-être (*welfare contractualism*). Il lui permet également de mettre l'accent sur la dimension hédoniste des raisons que Scanlon tend, selon lui, à minimiser. Reste que le modèle des raisons déontiques et non-déontiques développé par Parfit ne peut être présenté comme un conséquentialisme de la règle dans la mesure où l'évaluation de la force des raisons est fonction des personnes auxquelles elles s'imposent. Il ne peut donc pas y avoir de systématisme de la règle dans le modèle de Parfit.

### 3. Ethique appliquée

Les questions éthiques qui se posent à nous aujourd'hui sont nombreuses. L'évolution technologique, de plus en plus rapide, associée à la mutation des normes sociales, soulève un certain nombre de controverses morales qui doivent être résolues<sup>224</sup>. Parmi les plus actuels, citons la bioéthique<sup>225</sup>, l'éthique climatique<sup>226</sup>, l'éthique des robots<sup>227</sup>, l'éthique de

---

<sup>223</sup> D. Parfit, *On What Matters*, op.cit., vol.2., pp.549-565.

<sup>224</sup> W. Shaw, *Contemporary Ethics: Taking Account of Utilitarianism*, Wiley-Blackwell, Londres, 1999.

<sup>225</sup> J. Glover, J., *Questions de vie ou de mort*, traduction et introduction de Benoît Basse, Labor et Fides, Genève, 2017 ; P. Singer, *Practical Ethics*, Cambridge: Cambridge University Press, 1993.

<sup>226</sup> J. Broome, *Climate matters. Ethics in a warming world*, New York, London:W.W. Norton & Company, 2012.

<sup>227</sup> M. Anderson et S. L. Anderson (eds), *Machine Ethics*, Cambridge: Cambridge University Press, 2011.



la guerre<sup>228</sup>, l'éthique des affaires<sup>229</sup> ou encore l'éthique de la population<sup>230</sup>. Bien que prenant des formes nouvelles, un grand nombre de ces débats se réduisent *in fine* à des débats éthiques plus anciens portant sur les fins de la médecine et de la technologie, la perfectibilité de l'homme ou encore le fondement du naturalisme moral. On pourrait donc arguer que les théories éthiques existantes suffisent à nourrir, sinon trancher, les débats. Il existe, cependant, au moins, une grande question éthique qu'aucune grande philosophie morale n'a su anticiper, celle du réchauffement climatique et de ses conséquences probables sur la biodiversité et le futur de l'espèce humaine. La possibilité, à plus à moins long terme de l'extinction de l'espèce humaine, ainsi que de nombreuses autres espèces animales, et la certitude, qu'à court et moyen terme, les générations futures auront à subir les conséquences dramatiques du dérèglement climatique ainsi que de la dégradation des écosystèmes, constitue sans aucun doute une question éthique de nouvel ordre, et peut-être la plus fondamentale d'entre toutes.

#### A) ETHIQUE APPLIQUEE ET THEORIE NORMATIVE

Le développement de l'éthique appliquée dans les années 1970, sous l'impulsion de Peter Singer et de Jonathan Glover notamment, a beaucoup contribué à son renouvellement. Je suis d'avis qu'une nouvelle vague de travaux sur le réchauffement climatique, le véganisme

---

<sup>228</sup> W. Shaw, *Utilitarianism and the Ethics of War*, Londres: Routledge, 2016.

<sup>229</sup> A. Gustafson, "In Defense of a Utilitarian Business Ethic", *Business and Society Review*, 118 (3), 2013, pp.325-360.

<sup>230</sup> G. Arrhenius, J. Ryberg et T. Tännsjö, "The Repugnant Conclusion", *The Stanford Encyclopedia of Philosophy*, in Zalta E.N. (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/repugnant-conclusion/>, 2017.

ou la grande pauvreté, vont dans un futur proche nous amener à reconsidérer nos positions théoriques, notamment à l'égard de l'utilitarisme de la règle.

En tant que philosophie morale, l'utilitarisme se doit de prendre part aux débats moraux qui secouent notre société. Dans les années 1970 et 80, les philosophes ont été appelés à se prononcer sur la légitimité morale de l'avortement, de l'euthanasie ou de la guerre. Aujourd'hui, ils sont consultés sur les politiques à adopter face au réchauffement climatique, la souffrance ou l'exploitation animale ou encore la procréation assistée (GPA-PMA). Certaines de ces questions exigent parfois d'élargir ou de modifier le cadre théorique de référence, par exemple en incluant/renforçant la dimension intergénérationnelle et/ou antispéciste dans le modèle de base. L'évolution de la théorie normative peut également être le produit de débats ayant lieu en amont, au niveau de la théorie méta-éthique<sup>231</sup> [Tännsjö, 1976].

La plupart des travaux d'éthique appliquée sont totalement déconnectés de l'éthique normative, qui devrait pourtant leur assurer un socle théorique. Cette déconnection est parfois même explicitement assumée. Ainsi, Tony Hope, dans son *Medical Ethics. A Very Short Introduction* (2004), explique que « dans ce livre, je n'essaie pas d'approcher les différents problèmes que je discute du point de vue d'une seule théorie morale. Chaque chapitre traite d'une question pour laquelle je défends une position particulière, *en utilisant les méthodes de raisonnement qui me semblent les plus appropriées* »<sup>232</sup>. La limite de ce type de posture est naturellement de permettre toutes les justifications possibles. Si

---

<sup>231</sup> T. Tännsjö, *The Relevance of Metaethics to Ethics*, Stockholm: Almqvist & Wiksell International, 1976.

<sup>232</sup> T. Hope, *Medical Ethics. A Very Short Introduction*, Oxford: Oxford University Press, 2004, p.6. Ma traduction, je souligne.

on est plutôt opposé à la Gestation pour Autrui, on a plutôt recours à la déontologie kantienne permet de justifier et si je suis favorable à l'avortement, je fais plutôt appel à l'utilitarisme. C'est, de mon point de vue, la pire des postures à adopter. Les justifications éthiques doivent être cohérentes entre elles. En pratique, cela pose néanmoins certains problèmes. Ainsi que je l'ai montré dans le cas de la Gestation pour Autrui<sup>233</sup>, les théories éthiques normatives existantes ont parfois du mal à traiter des questions nouvelles. On peut, certes, essayer de les interpréter, voire d'interpréter l'intention de l'auteur, mais les résultats sont souvent discutables. Nombreux sont ceux, par exemple, qui ont abusivement recours à Kant pour condamner une telle pratique en arguant du fait que la déontologie kantienne interdit l'instrumentalisation du corps humain. C'est oublier, cependant, que lorsqu'elle a pour objet la procréation, l'utilisation du corps humain (par exemple dans l'acte sexuel) est légitime. Kant, en revanche, ne dit rien de la procréation non sexuelle. Lui faire tenir de tels propos relève de la ventriloquie intellectuelle.

J'ai tenté dans un article sur la Gestation pour Autrui (*Raison Publique*, 2019), de montrer que contrairement aux idées reçues, les théories éthiques à notre disposition (conséquentialisme et déontologie) ne permettaient pas de répondre de façon satisfaisante aux demandes de la société. Les conclusions du CCNE, en plus d'être fortement variables d'années en années, sont faiblement justifiées et expriment plus souvent qu'elles ne le devraient une position dictée par la déontologie médicale et non par la morale. Les états généraux de la bioéthique, menés au premier semestre 2018, n'ont pas été davantage capables de fournir de réponses aux questions posées. Au mieux ont-elles été en mesure de mesurer l'ampleur des désaccords de nos concitoyens sur ces sujets. Il

---

<sup>233</sup> C. Salvat, « L'éthique face à la GPA », *Raison Publique*, n°23, 2019, pp.163-173.

est nécessaire que les philosophes s’emparent de ces questions, de toutes, et si possibles de façon simultanée. Les questions d’éthique appliquée ne peuvent, à mon sens, être traitées indépendamment les unes des autres, comme c’est trop souvent le cas. Elles doivent s’inscrire dans un cadre théorique commun, lui-même fondé sur des bases méta-éthiques cohérentes et solides. Là est la vraie difficulté à laquelle les philosophes doivent faire face. Les nouveaux enjeux de l’éthique, que ce soient l’IA, l’Homme Augmenté, la bioéthique ou l’extinction des espèces vivantes, n’appellent pas uniquement des réponses politiques, ils appellent avant tout des réponses théoriques. Je n’ai naturellement pas l’ambition de produire une théorie capable de répondre à la fois à tous ces enjeux mais, au moins, de travailler dans ce sens. Ainsi que je l’ai précédemment expliqué, je crois que le travail de Parfit constitue une excellente base théorique. Je crois, notamment, que sa distinction entre raison déontique et raison non-déontique a de forts potentiels théoriques voire pratiques. Je propose donc de poursuivre une réflexion éthique, autant dans sa dimension théorique qu’appliquée, à l’aide des outils conceptuels fournis.

En matière d’éthique appliquée, mes travaux portent essentiellement sur les nouveaux enjeux de l’éthique, autrement dit sur les questions soulevées par l’actualité (la procréation assistée, l’IA, les Big Datas et le réchauffement climatique), questions que les théories traditionnelles ont parfois du mal à prendre en compte. J’ai eu l’occasion de présenter mes premières pistes de réflexion au cours d’un cycle de conférences données à l’Université du Temps Libre d’Aix-en-Provence en 2018-19, cycle que je poursuis en 2019-2020. J’attache une importance particulière à la question du réchauffement climatique, ou plus exactement des réponses morales que l’on peut lui donner.

## B) L'ETHIQUE ANIMALE

Le principe de l'impartialité nous enjoint à ne jamais prendre personnellement parti dans une délibération morale et à considérer toutes les parties affectées comme ayant un statut égal. Selon l'expression de Bentham, « chacun compte pour un et uniquement pour un ». Un certain nombre de penseurs, dont Bentham lui-même, se sont cependant posé la question de savoir si cette impartialité ne s'appliquait pas également à *toutes* les parties affectées, y compris les animaux<sup>234</sup>. Dans une note de son *Introduction aux principes de morale et de législation*, Bentham remarque ainsi que le critère hédoniste retenu implique que les parties affectées par une action sont logiquement celles qui sont en capacité d'en souffrir. Or la souffrance n'est pas une spécificité humaine : « Pourquoi [leurs intérêts] ne sont-ils pas, universellement, tout autant que ceux des créatures humaines, considérés en fonction des différences de degré de sensibilité ? [...] Autrefois, et j'ai peine à dire qu'en de nombreux endroits cela ne fait pas encore partie du passé, la majeure partie des espèces, rangée sous la dénomination d'esclaves, étaient traitées par la loi exactement sur le même pied que, aujourd'hui encore, en Angleterre par exemple, les races inférieures d'animaux. Le jour viendra peut-être où il sera possible au reste de la création animale d'acquérir ces droits qui n'auraient jamais pu lui être refusés sinon par la main de la tyrannie.[...] La question n'est pas « peuvent-ils raisonner? », ni « peuvent-ils parler ? », mais « peuvent-ils souffrir? ». »<sup>235</sup>.

Longtemps restée secondaire, cependant, l'intégration de la souffrance animale dans le calcul de maximisation a été sérieusement prise en compte à partir des années 1970, à

---

<sup>234</sup> J. Kniess, "Bentham on animal welfare", *British Journal for the History of Philosophy*, vol.27, n°3, 2019, pp.556-572.

<sup>235</sup> J. Bentham, *Introduction aux principes de morale et de législation* [1789], traduction par le centre Bentham, Paris : Vrin, 2011, pp.324-25.

nouveau sous l'impulsion de Peter Singer. Parallèlement à sa lutte contre la pauvreté et les famines dans le monde, Peter Singer s'est en effet fait connaître pour son engagement précoce pour la cause animale. Son premier livre, *Animal Liberation*, initialement publié en 1975, a joué un rôle moteur dans la lutte contre les expérimentations, l'exploitation et la consommation des animaux. Ses travaux ont eu une influence importante et immédiate. Inspiré par Singer, l'activiste Henry Spira a ainsi réussi à mobiliser l'opinion publique américaine contre l'expérimentation sur les animaux dans l'industrie cosmétique, en menant, dans les années 1970 et 1980, des campagnes publicitaires très préjudiciables pour les marques<sup>236</sup>.

L'argument de Singer repose sur l'analogie entre la discrimination sexuelle ou raciale et la discrimination spéciste. Le spécisme (il est un des premiers à employer et diffuser cette expression) est, pour lui, tout aussi condamnable que le sexisme ou le racisme car il viole le principe d'impartialité : « Le spécisme est un préjugé ou une attitude de parti pris en faveur des intérêts de sa propre espèce et à l'encontre des intérêts des membres des autres espèces »<sup>237</sup>. En plus de l'abandon (ou de la stricte limitation) des expérimentations animales, Singer exhorte ses lecteurs à prendre conscience de leur responsabilité, en tant que consommateur, dans la maltraitance des animaux, et, par conséquent, à cesser de consommer des produits d'origine animale.

Ce type de défense de la cause animale (il est difficile de parler de droits des animaux dans une perspective utilitariste) fait généralement l'objet d'une double attaque : elle est, tout

---

<sup>236</sup> P. Singer, *Théorie du tube de dentifrice. Comment changer le monde selon Henry Spira*, traduction par A. Pons, Paris : Editions Goutte d'Or, 2018.

<sup>237</sup> P. Singer, *La libération animale*, traduit par L. Rousselle, Paris : Petite Bibliothèque Payot, 2012, p.73.

d'abord, remise en cause par ceux qui considèrent qu'elle va 'trop loin' (en exigeant des hommes qu'ils changent leurs habitudes alimentaires, par exemple), mais est aussi fortement critiquée par ceux qui considèrent qu'elle ne va pas 'assez loin', en refusant d'accorder aux animaux de véritables droits.

L'utilitarisme est, en effet, un instrument à double tranchant pour les militants de la cause animale. S'il permet de prendre en compte la souffrance des animaux, ou leur préférence à rester en vie, il peut également justifier ces mêmes souffrances (ou leur mise à mort) au nom de l'utilité totale. Singer a ainsi été attaqué pour ne pas systématiquement rejeter les expérimentations animales, notamment celles considérées comme nécessaires pour la mise au point de traitements médicaux pour l'homme<sup>238</sup>. L'utilitarisme permet également de justifier la consommation et l'exploitation animale lorsque les conditions d'élevage des animaux sont jugées compatibles avec le bien-être animal. Selon l'argument dit du remplacement, des millions de vies animales – et l'utilité qui leur est associée – n'existeraient simplement sans l'élevage. Lorsque ces animaux sont tués, ils sont remplacés par d'autres. Richard Hare a proposé de recourir à l'indicateur QALY (*Quality-Adjusted Life Year*), généralement utilisé en médecine pour évaluer l'utilité d'une intervention, pour juger de la légitimité du régime végétarien. Selon lui, l'objectif de la maximisation QALY est difficilement compatible avec l'abandon complet de la consommation animale<sup>239</sup>. L'adoption d'un régime 'demi-végétarien' semble plus adapté. Il est également plus réaliste car moins exigeant.

---

<sup>238</sup> L. Gruen, "Utiliser la Philosophie pour Changer le Monde", *Klesis*, n°32, 2016, pp.15-39.

<sup>239</sup> R.M. Hare, "Why I Am Only a Demi-vegetarian", in Jamieson, D., *Singer and his Critics*, Oxford: Blackwell Publishing, 1999, p.240.

Singer refuse 'empiriquement' l'argument du remplacement, même si ses écrits ne permettent pas vraiment de le remettre théoriquement en cause<sup>240</sup> [Delon, 2016]. L'utilitarisme, on l'a vu, est multiforme et peut conduire à des résultats très différents selon le modèle théorique, la métrique ou le type d'agrégation adoptés. Prendre en compte le bien-être animal ne signifie pas, *a priori*, de pondérer également les préférences humaines et les préférences animales<sup>241</sup>. Une version 'modérée' de l'utilitarisme (*satisficing*) pourrait également se satisfaire d'une réduction des souffrances animales sans nécessairement remettre en cause le système d'élevage intensif. Enfin, selon la théorie du code moral idéal, pour être optimifque, une règle morale doit pouvoir être internalisée relativement aisément. Plus la règle est contraignante, plus son enseignement est coûteux, moins efficace elle est.

Certains jugeront que la flexibilité de l'utilitarisme constitue l'une de ses principales limites, d'autres y verront un atout. La souffrance animale n'est, cependant, pas la seule considération à prendre en compte dans le calcul utilitariste. L'exploitation animale a de nombreuses autres conséquences sur l'utilité totale. Selon l'Organisation des Nations unies pour l'alimentation et l'agriculture (FAO), environ un quart des terres émergées est mis en pâturage et un tiers de la production agricole sert à l'alimentation animale<sup>242</sup>. Au total, plus de deux tiers des terres agricoles seraient directement ou indirectement utilisées pour la production animale. Dans le même temps, environ huit cents millions de personnes souffrent de malnutrition dans le monde et plus encore ont des difficultés à acheter des céréales, dont le cours est maintenu artificiellement élevé par la demande liée à

---

<sup>240</sup> N. Delon, « Un Singer peut-il en remplacer un autre ? », *Klesis*, n°32, 2016, pp.150-190.

<sup>241</sup> S. Kagan, *How to Count Animals, more or less*, Oxford: Oxford University Press, 2019.

<sup>242</sup> <http://www.fao.org/3/ar591e/ar591e.pdf>



l'alimentation animale. Le recours à l'agriculture intensive et à la déforestation massive pour augmenter la production céréalière a, de son côté, d'importantes conséquences négatives sur l'environnement. Il est donc relativement aisé de démontrer, d'un point de vue utilitariste, qu'une réduction, voire l'abandon total, de la consommation animale profiterait au bien-être des animaux comme des humains. Toute la difficulté, cependant, est de savoir si cela suffit pour constituer une obligation morale.

### C) LE RECHAUFFEMENT CLIMATIQUE

*Last but not least*, les théories utilitaristes nous fournissent les outils conceptuels pour nous interroger sur notre devoir moral face à la réalité et aux conséquences probables de nos actions sur le réchauffement climatique. Dans son rapport de 2019, le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) insiste sur la nécessité de respecter les objectifs d'émission de gaz à effets de serre approuvés lors de l'Accord de Paris de 2015. Les experts mettent notamment les gouvernements en garde contre la fonte des surfaces glacées, l'élévation du niveau des océans, et le bouleversement des écosystèmes terrestres et marins anticipés. Certaines zones géographiques pourront être submergées, d'autres pourraient se désertifier, d'autres enfin feront l'objet de catastrophes climatiques (tornades, cyclones, canicules etc.) de façon plus régulière et plus intense. Toutes les populations verront leurs conditions de vie impactées, mais certaines, en général celles des pays du Sud, seront plus durement affectées que les autres.

La dimension collective de la catastrophe annoncée semble a priori constituer un facteur favorable au principe d'impartialité. Certains éléments cités ci-dessus en compromettent l'exécution. Certaines régions du monde devraient, tout d'abord, être plus affectées que

d'autres. Les générations futures seront, ensuite, plus durement frappées que les générations présentes. Enfin, toutes les espèces animales ne seront pas menacées de la même manière. Les humains de la présente génération et vivants dans les pays industrialisés du Nord seront, en moyenne, mieux protégés que les autres. Ce sont, pourtant, les seuls en position de contrer le phénomène. Une remise en cause du principe d'impartialité peut, dès lors, avoir des effets irréversibles et dramatiques.

Au-delà de la question de l'impartialité, l'utilitarisme pose un certain nombre de questions auxquelles nous devons répondre si nous voulons nous engager dans une réflexion morale rigoureuse sur le futur de l'humanité : Devons-nous faire le *maximum* pour éviter le scénario avancé, au risque de sacrifier l'essentiel de notre confort de vie, ou pouvons-nous nous contenter d'un effort louable (*utilitarisme modéré*) ? Avons-nous réellement une obligation de maximiser le bien-être des générations futures ou pouvons-nous nous contenter de minimiser leurs souffrances à venir (*utilitarisme négatif*) ? Devons-nous calculer l'utilité des futurs agents en nous fondant sur nos préférences actuelles ou ce que pourraient être leurs préférences dans un lointain futur (*utilitarisme des préférences*) ? Devons-nous assumer une responsabilité morale individuelle face à cette sombre perspective ou pouvons-nous répartir les efforts à consentir entre tous (*utilitarisme indirect*) ? Devons-nous établir un code moral, facilement assimilable, qui s'applique à tous tout le temps, plutôt que de laisser les individus délibérer sur la légitimité de leurs actions (*utilitarisme de la règle*) ? Devons-nous répartir les sacrifices de façon inégale pour éviter que les plus pauvres paient un prix disproportionné pour une situation dont ils ne sont que très faiblement responsables ?

Ces travaux constituent la première étape d'un programme de recherche plus vaste sur l'articulation des différents niveaux de l'éthique, et sur la nécessité de ne pas sous-estimer l'importance de l'éthique appliquée comme de la méta-éthique en éthique normative.

## QUATRIEME PARTIE

### ANIMATION DE LA RECHERCHE ET CV

## 1. Animation de la recherche

J'ai soutenu une thèse d'histoire de la pensée économique au Centre Auguste et Léon Walras de Lyon en 2000. J'ai ensuite passé un an à l'Université de Lausanne en tant que maître-assistant à la faculté de droit, puis quatre ans à Aix-en-Provence en tant que chargé de recherche, dans un laboratoire d'économie (le Greqam, aujourd'hui AMSE), cinq ans à l'Université de Cambridge en tant que Research Fellow et Associate Lecturer à la faculté d'histoire, quatre ans à Lyon en tant que chercheur à Triangle en tant que coordinateur de l'axe Philosophie économique. Aujourd'hui, de retour à Aix-en-Provence, en tant que chercheur au Centre Gilles-Gaston Granger, je sollicite l'obtention d'une Habilitation à Diriger les Recherches.

Je considère cette HDR à la fois comme une reconnaissance du travail de recherche effectué jusqu'ici et comme une opportunité de les développer avec plus de moyens, notamment grâce à la possibilité que celle-ci donne d'encadrer des étudiants en thèse de doctorat. J'ai essayé de montrer ci-dessus que mes recherches avaient substantiellement évoluées depuis ma thèse, et que la plupart des travaux présentés dans le cadre de cette HDR étaient originaux. J'ai également essayé de montrer qu'en dépit de leur variété, ces travaux répondaient à une même ligne directrice, et que je considérais l'histoire de la pensée et l'analyse théorique comme deux approches complémentaires. De la même façon, ces travaux illustrent tous la dimension fondamentalement politique et morale du discours et des concepts économiques, justifiant l'approche interdisciplinaire qui a été la mienne depuis mes débuts. Je voudrais maintenant donner quelques éléments relatifs à mon expérience d'animation de la recherche ainsi que quelques éléments relatifs à mon programme de recherche pour les années à venir.

Depuis ma nomination au CNRS, je me suis beaucoup engagé dans la vie des laboratoires dans lesquels j'ai successivement travaillé.

J'ai coordonné le séminaire de philosophie économique du Greqam de 2002 à 2006 et celui de Triangle de 2013 à 2017. Ces séminaires réguliers (environ 8 fois par an pour le Greqam et 4 fois par an pour Triangle) étaient destinés aux étudiants de master et de doctorat ainsi qu'aux membres du laboratoire intéressés. Les intervenants étaient tous extérieurs, souvent étrangers.

J'ai également organisé quatre colloques internationaux. Le premier, intitulé *Men and Citizens: Nature, Foundations and Challenges of Citizenship*, organisé à Cambridge (UK) les 14 et 15 juillet 2008, en partenariat avec Robinson College, portait sur la notion de citoyenneté et rassemblait des historiens, des philosophes et des économistes. Il a donné lieu à la publication d'un numéro spécial de *History of European Ideas* (Volume 37, Issue 2, June 2011). Le second, intitulé *Paternalism Redeemed: Old Ideals, New Realities*, a eu lieu à l'Ecole Normale Supérieure de Lyon les 30 et 31 mars 2015, grâce à un financement Projet émergent du fonds recherche de l'ENS de Lyon. Le troisième, intitulé *Paternalism, Liberalism, Socialism : France and Britain, 19th century*, co-organisé avec Ludovic Frobert, a eu lieu à la Maison Française d'Oxford (Oxford, UK) les 3 et 4 juin 2016. Il a été co-financé par le programme ANR Saint-Simonisme 18-21.

Membre du réseau de philosophie économique, j'ai co-organisé avec Antoinette Baujard et Charles Girard le 4ème Colloque International de Philosophie Economique, qui a eu lieu du 27 juin 2018 au 29 juin 2018 à l'ENS de Lyon et qui avait pour thème « Normes et

Normativités ». Il a accueilli près de deux cents participants dont les deux tiers étrangers. Enfin j'ai participé à l'organisation du 15e Colloque international de l'Association Charles Gide pour l'étude de la pensée économique (sur le thème « Les économistes et le pouvoir») qui a eu lieu les 2, 3 et 4 juin 2014 à Lyon.

Au niveau de la gouvernance du laboratoire, j'ai été membre élu du conseil de laboratoire du GREQAM de 2003 jusqu'à 2006, membre du Governing Body de Robinson College (Cambridge, UK) de 2006 à 2011 et membre nommé du conseil de laboratoire de Triangle de 2015 à 2018. Dans ce dernier laboratoire, j'ai été responsable de l'axe de Philosophie économique de 2015 à 2018. Je suis aujourd'hui directeur adjoint de l'axe 2 (Histoire de la philosophie et ontologies du présent) du Centre Gilles-Gaston Granger.

Par ailleurs, je suis membre élu du conseil d'administration de l'Association Charles Gide pour l'Etude de la Pensée Economique depuis 2016. Je l'ai également été de 2003 à 2008, période pendant laquelle j'ai assuré le poste de secrétaire général (2004-2006).

Je suis membre du comité de rédaction de '*Economie & Société, série PE*', depuis 2011 devenue *Revue d'Histoire de la Pensée Economique* en 2016 (Classiques Garnier). J'ai également coordonné plusieurs numéros spéciaux de revues, dont « Trente ans d'histoire de la pensée économique, 1984-2014 », *Economies et Sociétés*, série PE, n°51, décembre 2014, le numéro 5 de la *Revue d'Histoire de la Pensée Economique* (n°1, janvier 2018), le volume 37, « Rethinking Citizenship » (n°2) de *History of European Ideas*, special issue " (June 2011) ainsi que, (en collaboration avec Laurent Jaffro), le numéro spécial de la *Revue de Métaphysique et de Morale* consacré à Parfit (2019, n°2).

J'ai enfin donné de nombreux enseignements en économie, en histoire de la pensée politique (à l'université de Cambridge, où j'ai été Associate Lecturer de 2009 à 2011) et en philosophie économique (à l'Université de Provence et à l'ENS de Lyon). Je donne également des conférences et séminaires tout public. J'ai ainsi donné cette année un cycle de six conférences consacrées aux nouveaux enjeux de l'éthique à l'Université de Temps Libre (AMU), accompagnées de séminaires, ainsi qu'un certain nombre de conférences publiques (Bibliothèque de l'Alcazar, Bordeaux Geek Festival, conférences ciné-débat etc..). Ayant enseigné en Master 2 depuis ma nomination au CNRS, j'ai eu l'occasion d'encadrer un grand nombre d'étudiants dans leurs travaux de recherche. Je souhaite aujourd'hui poursuivre ce travail et pouvoir les accompagner tout au long de leur thèse. Je co-encadre déjà l'un de mes anciens étudiants de l'ENS, David Philippy, aujourd'hui en thèse de doctorat à l'université de Lausanne sous la direction de Harro Maas. Je travaille actuellement avec Izem Dahmani, qui doit se lancer très prochainement dans la rédaction d'une thèse sur le transhumanisme, et qui souhaite que je puisse l'encadrer officiellement.





## 2. Curriculum Vitae

### FONCTIONS ET STATUTS

- **Mars 2018 - :** Chargé de recherche, Centre Gilles Gaston Granger UMR 7304, Université d'Aix-Marseille.
- **Mai 2013- Février 2018:** Chargé de recherche CNRS, Triangle UMR 5206, Ecole Normale Supérieure de Lyon.
- **Octobre 2002- Mai 2013:** Chargé de recherche CNRS, GREQAM, UMR 65-79, Aix-en-Provence
- **Octobre 2009-2011:** Associate Lecturer, University of Cambridge
- **Octobre 2006- 2011:** Fellow, Robinson College, Cambridge, UK.
- **Septembre 2001-Septembre 2002 :** Maître-assistant à l'Université de Lausanne (Suisse)
- **Septembre 2000-Septembre 2001 :** ATER à l'Université Lumière Lyon II
- **Septembre 1999-Septembre 2000 :** ATER à l'Université Lumière Lyon II
- **Décembre 1996-Septembre 1999 :** Allocataire de recherche ; Doctorant au Centre Auguste et Léon Walras (Université Lumière, Lyon II) et à Trinity College (University of Dublin)

### INSTITUTIONS UNIVERSITAIRES DE RATTACHEMENT

- Centre Gilles-Gaston Granger, UMR7304, Aix-en-Provence.
- Triangle, UMR 5206, ENS Lyon, chargé de recherche 1ère classe CNRS.
- Robinson College, Cambridge (UK), Senior Member
- Centre for History and Economics, King's College, Cambridge (UK), associate fellow.

# TRAVAUX SCIENTIFIQUES

## Thèse de doctorat

- “Formation et diffusion de la pensée économique libérale française : André Morellet et l'économie politique du dix-huitième siècle”, soutenue le 14 janvier 2000 à l'Université Lumière Lyon II, 2 vols.

## Ouvrage

- *L'utilitarisme*, La Découverte, parution automne 2020.

## Articles publiés dans des revues à comité de lecture:

[19] “Still-Born Yet not Without Influence: What Mill’s Political Economy Owes to His Project of Ethology”, *Journal of the History of Economic Thought*, à paraître (2021).

[18] “Des personnes aux raisons : autour de Derek Parfit” (avec Laurent Jaffro), *Revue de Métaphysique et de Morale*, n°2, 2019, pp.3-6.

[17] “Parfit et les raisons du conséquentialisme de la règle”, *Revue de Métaphysique et de Morale*, n°2, 2019, pp.37-48.

[16] “L'éthique face à la GPA”, *Raison Publique*, n°23, 2019, pp.163-73.

[15] “French Political Economy and Positivism (or how history of economic thought became mainstream economics)”, *Iberian Journal of the History of Economic Thought*, vol4., n°2, 2017.  
<https://revistas.ucm.es/index.php/IJHE/article/view/58093>

[14] “From Literature to Politics: How Rousseau Has Come to Symbolize Totalitarianism”, *Journal of Interdisciplinary History of Ideas*, 2017, vol.6, n°11, pp.4-23.

[13] “Economics of paternalism: the hidden costs of self-commanding strategies”, *The Journal of Philosophical Economics: Reflections on Economic and Social Issues*, 9(1), 2015, pp.102-124.

- [12] “Parfit l'égoïsme rationnel et la question de l'identité personnelle”, *Oeconomia, History, Methodology, Philosophy*, 5(4), 2015, pp.437-60.
- [11] “Behavioral Paternalism”, *Revue de Philosophie Economique*, vol.15, n°2, 2014, pp.109-130.
- [10] “Rousseau et la « Renaissance classique » française (1898-1933) ”, *Astérion, Philosophie, Histoire des Idées, Pensée politique*, 12, 2014 [<http://asterion.revues.org/2545>]
- [9] “Metamorphoses of Citizenship”, *History of European Ideas*, 2011, vol.37, n°2, pp.209-210.
- [8] “Autonomie morale et autorité, Ou la question de la volonté chez Rousseau”, *Cahiers d'Economie Politique*, n°53, Décembre 2007, 75-93
- [7] “La loi et l'échange: le statut de la rationalité économique chez Rousseau”, *Revue économique*, volume 58, n°2, Mars 2007, pp.381-398.
- [6] “*Oeconomies*. Les articles OE/Economie et leurs désignants dans l'*Encyclopédie*”, *Recherches sur Diderot et l'Encyclopédie*, n°40, Octobre 2006, pp.97-116.
- [5] “Reason and Sentiments: Review of Emma Rothschild's *Economic Sentiments*. Adam Smith, Condorcet and the Enlightenment” (review article en collaboration avec Pascal Bridel), *European Journal of the History of Economic Thought*, vol.1, n°1, 2004, pp.131-145.
- [4] “'De la science de l'économie publique' de Morellet ”, *Studies on Voltaire and the Eighteenth Century*, October 2003, pp.65-96.
- [3] “La *Scottish Subtlety*. The commentaries of André Morellet on the *Wealth of Nations* ” en collaboration avec Richard van den Berg, *European Journal of the History of Economic Thought*, 2001, vol.8, n°2, pp.146-185.
- [2] “ André Morellet et la diffusion de la pensée économique italienne ”, *Il pensiero economico italiano*, Pise, VIII, 2000, n°2, pp.99-124.
- [1] “ Histoire de la traduction inédite de la *Richesse des nations* par l'abbé Morellet. "Une traduction manuscrite toujours célébrée et toujours obstinément refusée au public." ”, *Storia del Pensiero Economico*, Florence, 1999, n.38, pp.119-136.

## Contributions à des ouvrages collectifs ou encyclopédies

- [9] “Parfit”, in *Encyclopédie Philosophique*, (avec comité de lecture), janvier 2019, <http://encyclo-philo.fr/parfit-a-2/>.
- [8] “Egoïsme, altruisme et justice distributive”, in P. Savidan (dir.), *Dictionnaire des inégalités*, Paris, Puf, 2018, pp.433-447.
- [7] “Derek Parfit”, in P. Savidan (dir.), *Dictionnaire des inégalités*, Paris, Puf, 2018, pp.127-1221.
- [6] “Rousseau’s virtuous economics”, in R. Kuroki and Y. Ando (Eds.) *The Foundations of Political Economy and Social Reform: Economy and Society in Eighteenth Century France* (Routledge Studies in the History of Economics), London: Routledge, 2018, pp.134-154.
- [5] “Représenter la volonté générale: Rousseau et son héritage paradoxal” [with Théophile Pénigaud de Mourgues], in M. Albertone et D. Castiglione, *Les métamorphoses de la représentation. Langages et pratiques politiques à l'époque moderne*, Paris, Classiques Garnier, 2018, pp.115-137.
- [4] “L’ambivalence de la référence physiocratique chez Walras”, in *Études walrassiennes*. Roberto Baranzini, Arnaud Diemer et Claude Mouchot (éds.), Paris: L'Harmattan, 2004, pp. 29-50.
- [3] “ Histoire du syntagme et du concept de division du travail au XVIIIème siècle ”, *Dictionnaire des usages socio-politiques (1770-1815)*, Paris : Champion, 2003, pp.39-66.
- [2] “ L’arithmétique politique d’André Morellet ”, (en collaboration avec Pierre Crépel), in Thierry Martin (dir.), *L’arithmétique politique française*, Paris : INED, 2003, pp.195-229.
- [1] “ Morellet André (1727-1819) ”, in Frobert L., Potier J.P., Tiran A. (sous la direction de), *Economistes en Lyonnais, en Dauphiné et en Forez*, Lyon : Institut des Sciences de l’Homme, 2000, pp.228-240.

## Direction de revue

- [4] *Revue de Métaphysique et de Morale*, numéro spécial Parfit, 2019, n°2 (en collaboration avec Laurent Jaffro).
- [3] *Revue d'Histoire de la Pensée Economique*, n°5, Juin 2018.
- [2] *Trente ans d'histoire de la pensée économique, 1984-2014*, numéro spécial, *Economies et Sociétés*, série PE, n°51, décembre 2014 (parution janvier 2015)
- [1] *History of European Ideas*, special issue "Rethinking Citizenship" (Guest Editor), Volume 37, Issue 2, June 2011.

## Comptes-rendus d'ouvrages

- [7] "The Cosmos of duty. Henry Sidgwick's Methods of Ethics, by Roger Crisp", *Oeconomia, History, Methodology, Philosophy*, 8(4), 2018, pp.533-537.
- [6] "Living by Default". Book review of 'Choosing Not to Choose: Understanding the Value of Choice', by Cass R. Sunstein, *Journal of Economic Methodology*, 2017, vol.24, n°4, pp.430-434.
- [5] 'Reasons without Persons, by B. Hedden', *Economics and Philosophy*, 32(3), 2016, pp. 548-553.
- [4] 'Manipulation of Choice, Ethics and Libertarian Paternalism by Mark D. White', *Oeconomia, History, Methodology, Philosophy*, 5(4), 2015, pp.533-537.
- [3] "Mill et le paternalism", Compte-rendu de Gregory Claeys, *Mill and Paternalism*, Cambridge University Press, 2013, et de David O. Brink, *Mill's Progressive Principles*, Oxford University Press, 2013, *Economies et Sociétés*, série PE, numéro 51, 2014, pp.2045-2049.
- [2] "Eighteenth-Century Economics: Turgot, Beccaria and Smith and Their Contemporaries by Peter Groenewegen", *The Adam Smith Review*, volume 2, 2005, 275-279.
- [1] "Bonheur public et méthode géométrique par Marco Bianchini", *Revue de Philosophie économique*, Juin 2003, pp.143-150.

## Actes de colloques, articles de vulgarisation et divers

- [8] « La série "Black Mirror", sombre miroir des dérives de l'humanité », Interview pour Sud-Ouest, 19 mai 2018, <https://www.sudouest.fr/2018/05/19/la-serie-black-mirror-sombre-miroir-des-derives-de-l-humanite-5064716-4693.php>
- [7] « Entrevue avec Antoinette Baujard, Jean-Sébastien Gharbi, et Christophe Salvat », *Éthique et économique/Ethics and Economics*, 1 :4(1), 2017, pp.90-95.
- [6] « Pour une réglementation éthique de la GPA en France », *Libération*, 24 mai 2017 [[www.liberation.fr/.../pour-une-reglementation-ethique-de-la-gpa-en-france\\_1571737](http://www.liberation.fr/.../pour-une-reglementation-ethique-de-la-gpa-en-france_1571737)]
- [5] « Perfect Derek Parfit », *Libération*, 5 janvier 2017, [[www.liberation.fr/debats/2017/01/05/perfect-derek-parfit\\_1539091](http://www.liberation.fr/debats/2017/01/05/perfect-derek-parfit_1539091)]
- [4] « Quand les constructeurs automobiles relancent le débat philosophique », *Le Monde*, 28 octobre 2016.
- [3] « Rétrospective des articles publiés depuis 1984 », *Economies et Sociétés*, numéro spécial 'Trente ans d'histoire de la pensée économique', 51, 2014, pp. 2015-2043.
- [2] « 'On ne fait point d'industrie entre ciel et terre' : Walras, lecteur de Quesnay », *Cahiers du CERAS*, Hors série n°3, Mai 2004, pp.207-221.
- [1] « L'affaire de la Compagnie des Indes », *Revue européenne 2000*, Décembre 2003, pp.5-6.

## Communications dans colloques internationaux

- [30] “What can economics tell us about ethics? Lessons from the Moral Machine Experiment”, *Charles Gide Workshop 2019: Evolutions of the disciplinary boundaries of Economics with the other sciences*, Montreal (QC, Canada), 25–27 June 2019.
- [29] “What can economics tell us about ethics? Lessons from the Moral Machine Experiment”, Annual Conference of the European Society for the History of Economic Thought (ESHET), Lille, France 23-25 May 2019.
- [28] “Parfit et le conséquentialisme de la règle”, *17<sup>e</sup> colloque international de l'Association Charles Gide*, Nancy, 27-29 septembre 2018.

- [27] “Représenter la volonté générale: Rousseau et son héritage paradoxal” [with Théophile Pénigaud de Mourgues], *Le Metamorphosi Della Rappresentanza : Linguaggi Politici E Pratiche Di Governo Nell’Epoca Moderna*, Torino, 10-12 November 2016. [conférence invitée]
- [26] “Justifying the principle of temporal neutrality: A Synchronic approach”, *3rd International Conference in Economic Philosophy*, Aix-en-Provence, 15-16 June 2016.
- [25] “Mill and Paternalism : A ‘great misunderstanding’”, Conference on Paternalism, Liberalism, Socialism : France and Britain, 19th century, 3-4 juin 2016, Maison Française d’Oxford. [conference invitée]
- [24] “Paternalisme et économie : un duo gagnant ? ”, *16<sup>e</sup> Colloque International de l’Association Charles Gide pour l’Etude de la Pensée Economique*, 14-16 avril 2016, Strasbourg.
- [23] “Economics of Paternalism”, *Paternalism Redeemed : Old Ideals, New Realities*, Lyon 30th-31st March 2015, Ecole Normale Supérieure de Lyon. [conference invitée]
- [22] “‘Sur les causes qui nuisent à l’autorité de la science économique’ ou la crise de l’expertise économique sous le Second Empire”, *15<sup>e</sup> Colloque Charles Gide pour l’Histoire de la Pensée Economique*, Université de Lyon, 2-4 juin 2014.
- [21] “Prudential Foundations of Paternalism”, *2<sup>e</sup> Colloque International de ‘Philosophie économique’, Université de Strasbourg*, 9-10 Octobre 2014.
- [20] “Rousseau’s Epicurean Economics”, (Invited Speaker), International Conference “French Political Economy in the Age of Enlightenment”, Rikkyo University (Japan), 6-10<sup>th</sup> September 2012.[conference invitée]
- [19] “L’économie peut-elle rendre morale le paternalisme?”, *Colloque International de Philosophie Economique*, Science Po Lille, 21-22 juin 2012.
- [18] “Is Libertarian Paternalism an Oxymoron? A Comment on Sunstein and Thaler (2003)”, *XIIIth Annual Conference for the History of Economic Thought*, Prague, 15<sup>th</sup>-17<sup>th</sup> May 2008.
- [17] “Rationality, Empathy and Emotions: the Case of Compassion”, *Rationality and Emotions*, Cambridge University, 9-10 Janvier 2008.[conference invitée]
- [16] “Rousseau on Citizenship”, *History of Economic Thought Annual Conference*, Queen’s University, Belfast, 10-12 Septembre 2007.



- [15] “Individual Deliberation, Moral Autonomy and Ethics in Rousseau’s Thought”, *11th ESHET Conference*, Université Louis Pasteur – Strasbourg, 5-7 Juillet 2007.
- [14] “The status of the domestic model in Rousseau’s political economy”, *9th ESHET Conference "Economics and other disciplines"*, University of Stirling, 9-12 June 2005.
- [13] “L’économie politique, une branche nouvelle du savoir ?”, *Colloque international ‘Les branches du savoir dans l’Encyclopédie’*, Paris, 18-19 novembre 2004.[conférence invitée]
- [12] “La lance d’Achille ou l’économie politique de Rousseau”, *Journées en l’honneur de Jean Mathiot*, Aix-en-Provence, 30 septembre-2 octobre 2004.[conférence invitée]
- [11] “ L’influence de Vincent de Gournay dans l’Encyclopédie de Diderot et d’Alembert ”, *Colloque international "Commerce, population et société autour de Vincent de Gournay (1748-1758): La genèse d’un vocabulaire des sciences sociales en France"*, Paris, 19-21 février 2004.
- [10] “ De Forbonnais à Quesnay, les représentations de la concurrence dans l’Encyclopédie de Diderot et d’Alembert ”, *Xe colloque de l’Association Charles Gide pour l’Etude de la Pensée Economique*, Grenoble, 25-27 septembre 2003.
- [9] “ Un collaborateur zélé dans l’ombre de Turgot (1774-1776): l’abbé Morellet ”, *Colloque International "Turgot (1727-1781), notre contemporain? Economie, administration et gouvernement au Siècle des Lumières*, Université de Caen, 14-16 mai 2003.
- [8] “ ‘Industry and trade are not carried on in inter-stellar space’: Walras, as a reader of Quesnay”, *European Society of the History of Economic Thought Conference*, 15-17 mars 2002, Rethymno University, Crete.
- [7] “French macropolicy and S&T System : evolution and interlinkages (1981-1997) ” en collaboration avec Christian Lebas, *Macrotech Conference*, 1-3 mars 2002, Instituto de Estudios Internacionales, Madrid.
- [6] “ Morellet e l’Italia ”, *I reciproci influssi tra la Francia e l’Italia del XVIII secolo nel campo della scienza economica*, Université de Parme, Italie, 24-25 septembre 1999. [conférence invitée]
- [5] “ Le statut de l’économie politique chez André Morellet : pour la science, contre les mathématiques ”, *Colloque international de l’Association Charles Gide pour l’Etude de la Pensée Economique*, Université Paris I Panthéon-Sorbonne, 17-18 septembre 1999.

- [4] “ The French diffusion of *Della Moneta* from the abbot Ferdinando Galiani : the unrecognized part of André Morellet ”, *Tenth International Congress on the Enlightenment*, Dublin, Irlande, 25-31 juillet 1999.
- [3] “ Morellet et l’arithmétique politique anglaise : une critique inédite ”, *Colloque L’Arithmétique politique française*, Université de Besançon, 22-24 octobre 1998, 35p.
- [2] “Morellet’s theory of value. An original point of view in the French subjective theory of value ”, *Annual History of Economic Thought Conference*, Université de Bath, Angleterre, 7-9 Septembre 1998. [conference invitée]
- [1] “French translations of the *Wealth of Nations* in the eighteenth century: the case of the division of labour ”, *European Society for the History of Economic Thought Conference*, Université de Bologne, Italie, 27 février-1er mars 1998.

### **Conférences invitées et séminaires**

- [26] “Le Profane et l’Expert, ou de la déontologie du chercheur”, Conférence invitée, Doctoriales ED536 & ED537, Université d’Avignon, 13-14 juin 2019.
- [25] “La mort peut-elle être objet de désir ?”, 2<sup>e</sup> *Journées Phi Club*, Faculté des Arts, Lettres, Langues et Sciences Humaines, AMU, Aix-en-Provence, 11 avril 2019.
- [24] “Ethique sans méta-éthique n’est que ruine de la raison”, *Qu’est-ce que la méta-éthique ?*, IHP, AMU, Aix-en-Provence, 14 mars 2019.
- [23] “L’entreprise au secours de l’éthique ?”, *Secondes Journées Philosophie et Entreprises. Compétition, Concurrence, Coopération : Comment travailler ensemble dans le monde de demain ?*, IHP, AMU, Aix-en-Provence, décembre 2018.
- [22] “Parfit et l’utilitarisme des raisons”, Séminaire de l’AMSE, AMU, Marseille, 14 mars 2018.
- [21] “Normativity of Temporal Neutrality”, *Workshop Time and Action*, Oslo, 25-26 August 2016.
- [20] “The Possibility of Paternalism”, *Political Philosophy Workshop*, Faculty of Philosophy, Cambridge, 25th January 2013.
- [19] “Rousseau on Prudence”, *Cambridge Seminar in the History of Economic Analysis*, Clare Hall, Cambridge, 24<sup>th</sup> January 2013.

- [18] “Behavioral Paternalism”, *Séminaire du BETA*, Strasbourg, 23 Novembre 2012.
- [17] “Le Paternalisme Comportemental”, *Collège International de Philosophie*, Paris, 11 Janvier 2010.
- [16] “General Introduction: What is Citizenship?”, *Men and Citizens: Nature, Foundations and Challenges of Citizenship*, Cambridge, 14-15<sup>th</sup> July 2008.
- [15] “Individual Deliberation, General Will and Ethics”, 7th SCEME Seminar ‘Economics as a Moral Science’, Stirling University, 19 May 2007.
- [14] “L’échange et la loi : le statut de la rationalité économique chez Rousseau”, *Séminaire du CREUSET*, Université de Saint-Etienne, 18 Mai 2006.
- [13] “L’économie domestique de la *Nouvelle Héloïse*”, *Séminaire du CRAIRAC*, Aix-en-Provence, 25 mai 2005.
- [12] “Le statut du modèle domestique dans l’économie politique de Rousseau”, *Séminaire général du Centre Walras-Pareto*, Lausanne, 28 avril 2005.
- [11] “Autour de la *Lettre à d’Alembert* : l’économie politique de Jean-Jacques Rousseau”, *Séminaire général du Centre Walras-Pareto*, Lausanne, 1<sup>er</sup> avril 2004.
- [10] “L’économie politique dans l’*Encyclopédie* de Diderot et d’Alembert”, *Séminaire général du Centre Auguste et Léon Walras*, Lyon, 4 mars 2004.
- [9] “Le concept de concurrence au XVIIIe siècle”, *Séminaire du Centre Walras-Pareto*, 4 juillet 2003.
- [8] “L’économie politique dans l’*Encyclopédie*”, *Journées d’études d’Alembert*, CIRM (Luminy, France), 6 - 10 Janvier 2003.
- [7] “Walras, lecteur de Quesnay”, séminaire Centre Auguste et Léon Walras, Lyon, 14 février 2002, 21 p.
- [6] “La théorie physiocratique des prix”, séminaire *Histoire de la pensée économique*, Université de Lausanne (Suisse), 10 janvier 2002, 16 p.
- [5] “Prix et monnaie chez François Quesnay : quelques éléments d’appréciation”, séminaire *Les pensées monétaires dans l’histoire*, Centre Auguste et Léon Walras, 12 janvier 2001, 22p.
- [4] “La pensée monétaire de François Quesnay”, *Université d’été en histoire de la pensée et méthodologie économiques*, Université Lumière Lyon II- Gate, 4-9 septembre 2000, 25p.
- [3] “De *Valeurs et Monnaies* à la *Digression* de Morellet. Retour sur la première re-construction de la théorie de la valeur de Turgot”, *Université d’été en histoire de la*

*pensée et méthodologie économiques*, Université Louis Pasteur de Strasbourg, 6-11 septembre 1999, 35p.

[2] “ La traduction au 18ème siècle, l'exemple d'une traduction inédite de *Della Moneta* de Ferdinando Galiani ”, Centre Auguste et Léon Walras, 26 septembre 1998, 18p.

[1] “ La balance du commerce est-elle vraiment un concept mercantiliste ? : le point de vue d'André Morellet ”, Centre Auguste et Léon Walras, 23 juin 1997, 26p.

### **Conférences grand public, tables rondes**

[8] Cycle de conférences sur les nouveaux enjeux de l'éthique, Université du Temps Libre, Aix-en-Provence, octobre 2018-juin 2019 (six conférences).

[7] Table ronde, Journée d'étude interdisciplinaire sur le bien-être, Paris 1, PHARE, 6 avril 2019 [annulée pour cause de santé]

[6] “Éthique et Animalité”, Conférence publique, Bibliothèque de l'Alcazar, Marseille, 20 octobre 2018.

[5] “L'éthique des robots”, Fête de la science, Maison de la Recherche, AMU, Aix-en-Provence, 11 octobre 2018.

[4] “De l'Homo Sapiens à l'Homme Augmenté : Quel futur pour l'humanité ? ” Conférence publique, Bordeaux Geek Festival, 20 mai 2018.

[3] “Black Mirror, une série prophétique”, Bordeaux Geek Festival, Table ronde, 19 mai 2018.

[2] “ Black Mirror, « Bientôt de retour » ”, Conférence ciné-débat, Université Lyon 3, 20 mars 2018.

[1] “L'éthique des robots”, Talk du Mercredi, ENTPE, Lyon, 1er février 2017.

# ACTIVITES D'ENSEIGNEMENT

## HISTOIRE DES IDEES POLITIQUES, ECONOMIQUES ET PHILOSOPHIQUES

### Cours Magistraux

- Ethique, Université du Temps Libre, séminaire, (24h, 2018-19).
- Ethique professionnelle, Master sciences, Université Aix-Marseille (6h, 2018-19)
- Ethique, Master Ergologie, Université Aix-Marseille, (12h, 2018-2019).
- Ethique et justice sociale, Master These et master HPP Université Lyon 2 et ENS de Lyon (12h, 2016-2018)
- Ethique et justice sociale, Master These et master HPP Université Lyon 2 et ENS de Lyon (21h, 2015-2016)
- Philosophie économique, Master 2, Université Lyon 2, Ecole Normale Supérieure de Lyon, 12 heures (2012-2016)
- L'Utilitarisme, Master 2, Université Paul Cézanne, Aix-en-Provence, 10 heures (2003-2011)
- Utilitarianism and its Critics, 3rd year, Faculty of History and SPS, University of Cambridge, 4 hours, (2008-2011)
- Equality and Welfare, 3<sup>rd</sup> year, Faculty of History and SPS, University of Cambridge, 4 hours, (2010-2011)
- Ideas of Liberty, 3<sup>rd</sup> year, Faculty of History, SPS and Philosophy, University of Cambridge, 4 hours, (2009-2010)
- Political Thought and English Classical Political Economy, 2<sup>nd</sup> year, Faculty of History and SPS, University of Cambridge, 2 hours (2010-2011)

### Supervisions

- History of Political Thought, 1700-1890, 2<sup>nd</sup> and 3<sup>rd</sup> year SPS, University of Cambridge, (2007-2011)
- Political Philosophy and History of Political Thought 1890-today, 3<sup>rd</sup> year SPS and History, University of Cambridge, (2008-2011)

## **ECONOMIE**

### **Cours magistraux**

- Microeconomics, 1<sup>ère</sup> année, Université de Lausanne, Suisse (2001-2002)
- Macroeconomics, 1<sup>ère</sup> année, Université de Lausanne, Suisse (2001-2002)

### **Travaux dirigés**

- Fluctuations et Croissance économique, 3<sup>e</sup> année, Université de Lyon 2 (1997-2001)
- Economie monétaire, 3<sup>e</sup> année, Université de Lyon 2 (1997-2001)
- Economie internationale, 3<sup>e</sup> année, Université de Lyon 2 (1997-2001)
- Micro-économie, 2<sup>e</sup> année, Université de Lyon 2 (1999-2001)
- Macro-économie, 2<sup>e</sup> année, Université de Lyon 2 (1999-2001)

# RESPONSABILITES ADMINISTRATIVES

## Gestion de la recherche

- Co-responsable axe 2 CGGG
- Responsable de l'axe 'Philosophie économique' de Triangle, UMR5206, CNRS, Ecole Normale Supérieure de Lyon.
- Coordinateur du programme de recherche 'Prudence, Moralité et Paternalisme', Ecole Normale Supérieure de Lyon.

## Activités éditoriales

- Membre du comité de rédaction de 'Economie & Société, série PE' depuis 2009, aujourd'hui Revue d'Histoire de la Pensée Economique (Classiques Garnier)

## Organisations de colloques internationaux

- *4e Colloque International de Philosophie Economique : Normes et normativités*, ENS de Lyon, 27-29 juin 2018 (en collaboration avec A. Baujard et C. Girard).
- *Paternalism, Liberalism, Socialism: France and Britain, 19<sup>th</sup> century*, Maison Française d'Oxford, Oxford, 3-4 juin 2016 (en collaboration avec Ludovic Frobert)
- *Men and Citizens: Nature, Foundations and Challenges of Citizenship*, Cambridge, 14-15th July 2008
- *Paternalism Redeemed : Old Ideals, New Realities*, Lyon 30th-31st March 2015, Ecole Normale Supérieure de Lyon.

## Animation de séminaire

- Coordination du Séminaire de Philosophie économique, Ecole Normale Supérieure de Lyon, (2013-2017)

- Coordination du Séminaire de Philosophie économique, GREQAM, Université de Aix-en-Provence, (2002-2006).

### **Responsabilités Administratives diverses**

- Membre du comité de sélection MCF, Université de Strasbourg, 2018.
- Membre du Conseil d'administration de l'Association Charles Gide pour l'Etude de la Pensée Economique (2016-)
- Membre nommé du conseil de laboratoire de Triangle (UMR5206) (2015-2018)
- Member of the Governing Body, Robinson College, Cambridge, (2006-2011)
- Member of the Joint Liaison Committee, Robinson College, Cambridge, (2007-2011)
- Membre élu du conseil de laboratoire du GREQAM (CNRS), (2003-2006).
- Membre élu du conseil d'administration de l'Association Charles Gide pour l'Etude de la Pensée Economique, (2003-2008).
- Secrétaire général de l'Association Charles Gide pour l'Etude de la Pensée Economique, (2004-2006).
- Conseiller municipal de la commune de Dareizé (Rhône, France) (2001-2006).





# Bibliographie

- Adams, Robert Merrihew. 1976. "Motive Utilitarianism." *The Journal of Philosophy* 73 (14):467-481.
- Ainslie G. and Haslam N., 'Hyperbolic Discounting', in G. Loewenstein and J. Elster (eds.), *Choice over Time*, New York: Russell Sage Foundation, 1992, pp.57-92.
- Ainslie G., *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person*, Cambridge: Cambridge University Press, 1992.
- Akerlof G., 'Procrastination and Obedience', *The American Economic Review*, 81(2), 1991, pp.1-19.
- Anderson E., "What Is the Point of Equality?", *Ethics*, 109:2, 1999, pp. 287-337.
- Anderson M. et Anderson S. L. (eds), *Machine Ethics*, Cambridge: Cambridge University Press, 2011.
- Arendt H., *The Origins of Totalitarianism*, New York: Schocken Books, 1951.
- Arneson R., "Mill versus Paternalism" , *Ethics*, 90 (4), 1980, pp.470-489.
- Arneson, R., "Primary Goods Reconsidered." *Noûs* 24 (3), 1990, pp.429-454.
- Arrhenius, G., Ryberg, J. et Tännsjö, T., "The Repugnant Conclusion", *The Stanford Encyclopedia of Philosophy*, in Zalta E.N. (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/repugnant-conclusion/>, 2017.
- Arrow, K.J., *Social Choice and Individual Values*. New York: John Wiley & Sons, 1951.
- Ashraf N., Camerer C. F., Loewenstein G., "Adam Smith, Behavioral Economist", *Journal of Economic Perspectives*, vol. 19, no. 3, Summer 2005, pp. 131-145.
- Atkinson R. F., "J. S. Mill's "Proof" of the Principle of Utility", *Philosophy*, 32 (121), 1957, pp.158 – 167.
- Audard, C., P. Thierry, and J.S. Mill, *L'utilitarisme: Essai sur Bentham*, Paris : Presses Universitaires de France – PUF, 2012.
- Audard, C., *John Rawls*, London: Stocksfield Acumen Publishing Limited, 2007.

Ayer, A. J., "The Principle of Utility." In *Philosophical Essays*, London: Palgrave Macmillan UK, 1972, pp.250-270.

Backhouse, R., "Sidgwick, Marshall, and the Cambridge School of Economics ", *History of Political Economy*, 38 (1), 2006, pp.15-44.

Bain, A., *James Mill: A biography*, London: Longmans, Green, and co, 1882.

Ball T., 'The Formation of Character: Mill's 'Ethology' Reconsidered', *Polity*, 33(1), 2000, pp.25-48.

Baudeau N., *Première introduction à la philosophie économique: ou analyse des états policés*, Paris : Didot, 1771.

Baujard, A., "L'économie du bien-être est morte. Vive l'économie du bien-être !" In *Philosophie économique. Un état des lieux*, ed. G. Campagnolo and J.S. Gharbi, Paris: Editions Matériologiques, 2017, pp. 77-128.

Bénabou R. and Tirole J., 'Self-Confidence and Personal Motivation', *The Quarterly Journal of Economics*, 117(3), 2002, pp.871-915.

Bénabou R. and Tirole J., 'Willpower and Personal Rules', *Journal of Political Economy*, 112(4), 2004, pp.848-86.

Bentham, J.. *Oeuvres de Jérémie Bentham: essai sur la situation politique de l'Espagne, défense de l'usure. De l'organisation judiciaire et de la codification*, Bruxelles: Société Belge de Librairie, 1840.

Bentham, J., *Introduction aux principes de morale et de législation* [1789], traduction par le centre Bentham, Paris : Vrin, 2011.

Berlin I., « Two Concepts of Liberty », in I. Berlin, *Four Essays on Liberty*, Oxford: Oxford University Press, 1969, PP.118-72.

Blaug M., *Economic Theory in Retrospect*, Cambridge : Cambridge University Press, 1962-1997.

Bonitto, C., "L'utilitarisme idéal de Moore: un conséquentialisme non hédoniste." In *Deux siècles d'utilitarisme*, ed. Bozzo-Rey M. and E. Dardenne, Rennes: Presses Universitaires de Rennes, 2011, pp. 75-88.

- Bouveresse, J., "La théorie des fictions chez Bentham." In *Regards sur Bentham et l'utilitarisme*, ed. Kevin Mulligan and Robert Roth, Genève: Librairie Droz, 1993, pp. 87-98.
- Bradley, B., "Two Concepts of Intrinsic Value", *Ethical Theory and Moral Practice*, 9 (2), 2006, pp.111-130.
- Brandt, R.B., *A Theory of the Good and Right*, Oxford: Clarendon Press, 1979.
- Brink, D.O., "Mill's Ambivalence about Rights", *Boston University Law Review*, vol.90, 2010, pp.1669-1704.
- Brink, D.O., *Mill's Progressive Principles*, Oxford: Clarendon Press, 2013.
- Brink D.O., 'Prospects for Temporal Neutrality', in C. Callender, *The Oxford Handbook of Philosophy of Time*, Oxford and New York: Oxford University Press, 2011, pp. 353-81.
- Brink D.O., 'Rational Egoism and the Separateness of Persons', in J. Dancy, *Reading Parfit*, Oxford: Blackwell Publishers, 1997, pp.96-134.
- Brink D.O., 'Sidgwick and the rationale for rational egoism', in B. Schultz (ed.), *Essays on Henry Sidgwick*, Cambridge: Cambridge University Press, 1992, pp.199-240.
- Brink, D.O., "Prospects for Temporal Neutrality." In *The Oxford Handbook of Philosophy of Time* edited by Craig Callender, Oxford and New York: Oxford University Press, 2011, pp. 353-381.
- Brink, D.O., "Sidgwick's Dualism of Practical Reason." *Australasian Journal of Philosophy*, 66 (3), 1988, pp.291-307.
- Broad, C.D., *Five Types of Ethical Theory*, New York: Harcourt, Brace, 1930.
- Broad, C.D., "On the Function of False Hypotheses in Ethics," *International Journal of Ethics*, vol.26, n°3, 1916, pp.377-97.
- Broome, J., *Weighing Lives*, Oxford: Oxford University Press, 2006.
- Broome, J., "A cause of preference is not an object of preference", *Social Choice and Welfare*, 10 (1), 1993, pp.57-68.
- Broome, J., *Climate matters. Ethics in a warming world*, New York, London: W.W. Norton & Company, 2012.

- Burrow, J.W., *Whigs and liberals: continuity and change in English political thought*, Oxford Clarendon Press, 1988.
- Capaldi N., 'Mill's Forgotten Science of Ethology', *Social Theory and Practice*, 2(4), 1973, pp.409-420.
- Capaldi N., *John Stuart Mill: A Biography*, Cambridge: Cambridge University Press, 2004.
- Carpenter K.E, « Recherches sur la nature et les causes de la richesse des nations d'Adam Smith et politique culturelle en France », in *Economie et Sociétés*, Oeconomia, Histoire de la pensée économique, série P.E. n°24, 1995, pp.5-30.
- Charles, L., Lefebvre, F., Théré, C., *Le cercle de Vincent de Gournay: savoirs économiques et pratiques administratives en France au milieu du XVIIIe siècle*, Paris : Institut National des Etudes Démographiques, 2011.
- Cirillo, R., *The Economics of Vilfredo Pareto*, London: Taylor & Francis, 2012.
- Cléro, J.P. *Bentham: Philosophe de l'utilité*, Paris: Ellipses, 2006.
- Cobban A., *The Crisis of Civilization*, London: Jonathan Cape, 1941.
- Collini, S., *Public moralists: political thought and intellectual life in Britain, 1850-1930*, Oxford: Clarendon Press, 1991.
- Constant B., *De la liberté des anciens comparée à celle des modernes*, Paris: Fayard, 2010.
- Crisp, R., *Mill on Utilitarianism*, London: Routledge, 1997.
- Crisp R., *The Cosmos of Duty: Henry Sidgwick's Methods of Ethics*, Oxford: Oxford University Press, 2011.
- Crisp, R., *Routledge Philosophy GuideBook to Mill on Utilitarianism*, London: Taylor & Francis, 2002.
- Darnton, R., *L'aventure de l'Encyclopédie, 1775-1800: un best-seller au siècle des Lumières*. Translated by M.A. Revellat, Paris: Librairie académique Perrin, 1992.
- Darwall, S. L., "Pleasure as Ultimate Good in Sidgwick's Ethics", *The Monist*, 58 (3), 1974, pp.475-489.
- Davis, J.B., *Individuals and Identity in Economics*, Cambridge: Cambridge University Press, 2010.

- Davis, J.B., *The Theory of the Individual in Economics: Identity and Value*, London: Taylor & Francis, 2003.
- de Champs, E., *"La déontologie politique", ou, La pensée constitutionnelle de Jeremy Bentham*, Genève: Droz, 2008.
- De Lazari-Radek, K. et Singer P., *The Point of View of the Universe: Sidgwick & Contemporary Ethics*, Oxford: Oxford Univeristy Press, 2014.
- Delon, N., « Un Singer peut-il en remplacer un autre ? », *Klesis*, n°32, 2016, pp.150-190.
- de Waal F., *Primates and Philosophers: How Morality Evolved*, Princeton and Oxford, Princeton University Press, 2009.
- Dickens, C., *Hard Times*, London: HarperCollins Publishers Limited, 2012.
- Digeon C., *La crise allemande de la pensée française (1870-1914)*, Paris : Presses universitaires de France, 1959.
- Driver, J., "Global Utilitarianism", in B. Eggleston & D. Miller (Eds.), *The Cambridge Companion to Utilitarianism*, Cambridge: Cambridge University Press, 2014, pp.166-176.
- Dunn J., *The Political Thought of John Locke: An Historical Account of the Argument of the 'Two Treatises of Government'*, Cambridge: Cambridge University Press, 1982.
- Dworkin G., "Paternalism" In R. A. Wasserstrom (ed.), *Morality and the Law*, Belmont, California: Wadsworth Publishing Company, 1971, pp.107-26.
- Eco U., *Lector in fabula*, Paris : Grasset, 1985.
- Engel P., *Les vices du savoir. Essai d'éthique intellectuelle*, Marseille : Agone, 2019
- Elster J., *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge: Cambridge University Press, 1979.
- Feinberg J., "Legal Paternalism" , *Canadian Journal of Philosophy*, 1 (1), 1971, pp.106-24.
- Feldman F., *Pleasure and the Good Life*, New York-Oxford: Oxford University Press, 2004.
- Ferey S., 'Paternalisme libéral et pluralité du moi', *Revue économique*, 62(4), 2011, pp.737-50.

Fleurbaey, M. et P. Mongin, "Théorie du choix social et économie normative." In Dictionnaire de philosophie morale, ed. M. Canto-Sperber, Paris: Presses Universitaires de France, 1996, pp. 243-51.

Frankena, W. K., "Sidgwick and the Dualism of Practical Reason", *The Monist*, 58 (3), 1974, pp.449-467.

Frederick S., Loewenstein G. and O'Donoghue T., 'Time Discounting and Time Preference: A Critical Review' in G. Loewenstein, D. Read and R.F. Baumeister (eds.), *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, New York: Russell Sage Foundation, 2003, pp.13-86.

Gauthier, D.P., *Practical Reasoning: The Structure and Foundations of Prudential and Moral Arguments and Their Exemplifications in Discourse*, London: Clarendon Press, 1963.

Gilardone, M. "Contexte, sens et portée de l'approche par les capabilités de Amartya Kumar Sen: Vers une économie normative post-welfariste." PhD, Université Lyon 2, 2007.

Glover J., *Causing Death and Saving Lives*, London: Penguin Books Ltd, 1977.

Glover, J. *Questions de vie ou de mort*, traduction et introduction de Benoît Basse, Genève : Labor et Fides, 2017.

Godwin W., *An Enquiry Concerning Political Justice, and its Influence on General Virtue and Happiness*, London: G.G.J. and J. Robinson, 1793.

Greaves, H. et H. Lederman, "Extended Preferences and Interpersonal Comparisons of Well-being", *Philosophy and Phenomenological Research On line*, 2016.

Greaves, H., "A Reconsideration of the Harsanyi–Sen–Weymark Debate on Utilitarianism", *Utilitas*, 29 (2), 2016, pp.175-213.

Griffin J., *Well-Being: Its Meaning, Measurement and Moral Importance*, Oxford: Clarendon Press, 1986.

Gruen, L., "Must Utilitarians Be Impartial?" in Jamieson, D. (ed.), *Singer and his Critics*, Oxford: Blackwell Publishing, 1999, pp.129-149.

Gruen, L., "Utiliser la Philosophie pour Changer le Monde", *Klesis*, n°32, 2016, pp.15-39.

Gruzalski B., 'Parfit's Impact on Utilitarianism', *Ethics*, 96(4), 1986, pp. 760-83.

- Gul F. and Pensendorf W., 'Harmful Addiction', *The Review of Economic Studies*, 74(1), 2007, pp.147-72.
- Gustafson A., "In Defense of a Utilitarian Business Ethic", *Business and Society Review*, 118 (3), 2013, pp.325-360.
- Halévy, E., *La formation du radicalisme philosophique: La jeunesse de Bentham, 1776-1789*, Ed. M. Canto-Sperber: Presses universitaires de France, 1995.
- Hamburger, J., "James Mill on Universal Suffrage and the Middle Class", *The Journal of Politics*, 24 (1), 1962, pp.167-190.
- Hamburger, J., *Intellectuals in politics: John Stuart Mill and the philosophic radicals*, New Haven : Yale University Press, 1965.
- Hare, R.M., *Moral Thinking: Its Levels, Method, and Point*, Oxford: Oxford University Press, 1981.
- Hare, R.M., "Why I Am Only a Demi-vegetarian", in Jamieson, D., *Singer and his Critics*, Oxford: Blackwell Publishing, 1999, p.233-246.
- Harrison, R., *Bentham*, London: Routledge & Kegan Paul, 1983.
- Harrod, R. F., "Utilitarianism revised", *Mind*, vol.45, n°178, 1936, pp.137-156.
- Harsanyi, J. C., "Cardinal welfare, individualistic ethics and interpersonal comparisons of utility", *Journal of Political Economy*, (63), 1955, pp.309-21.
- Harsanyi, J. C. "Morality and the Theory of Rational Behavior", *Social Research*, 44 (4), 1977, pp.623-656.
- Hart, H.L.A., *Essays on Bentham: studies in jurisprudence and political theory*, Oxford: Clarendon Press, 1982.
- Haybron, D. M., "Mental State Approaches to Well-Being", In *The Oxford Handbook of Well-Being and Public Policy*, ed. by Matthew D. Adler and Marc Fleurbaey, Oxford: Oxford University Press, 2016, pp. 347-378.
- Herrstein R.J. and Prelec D., 'A Theory of Addiction', in G. Loewenstein and J. Elster (eds.), *Choice over Time*, New York: Russell Sage Foundation, 1992, pp.331-60.



- Hont I., *Jealousy of Trade: International Competition and the Nation-state in Historical Perspective*, Harvard: Harvard University Press, 2005.
- Hooker, B., *Ideal Code, Real World: A Rule-consequentialist Theory of Morality*, Oxford : Oxford University Press, 2000.
- Hope T., *Medical Ethics. A Very Short Introduction*, Oxford : Oxford University Press, 2004.
- Hurka, T., "Moore in the Middle", *Ethics*, 113 (3), 2003, pp.599-628.
- Hutchinson, B., *G. E. Moore's Ethical Theory: Resistance and Reconciliation*, Cambridge: Cambridge University Press, 2001.
- Itzkin, E.S., "Bentham's Chrestomathia: Utilitarian Legacy to English Education", *Journal of the History of Ideas*, 39 (2), 1978, pp.303-316.
- Jaffro L., «A qui faire porter le chapeau des valeurs ? Une discussion du fondamentalisme des raisons», *Revue de Métaphysique et de Morale*, 2019, 2, pp.51-64.
- Jonas H., *Le principe responsabilité*, Paris : Les Editions du Cerf, 1990.
- Kagan, S., *How to Count Animals, more or less*, Oxford: Oxford University Press, 2019.
- Kahneman D., Knetsch J. and R. Thaler, "Experimental Tests of the Endowment Effect and the Coase Theorem", *Journal of Political Economy*, vol. 98 (6), 1990, pp.1325-48.
- Kahneman, D. and A. Tversky, "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, 47 (2), 1979, pp.263-291.
- Kahneman, D., Wakker P., and R. Sarin, "Back to Bentham? Explorations of Experienced Utility", *The Quarterly Journal of Economics*, 112 (2), 1997, pp.375-406.
- Kandil, F., "La justice est aveugle: Rawls, Harsanyi et le voile d'ignorance", *Revue économique*, 65 (1), 2014, pp.97-124.
- Kaplow, L., "Primary Goods, Capabilities,... or Well-Being?", *The Philosophical Review*, 116 (4), 2007, pp.603-632.
- Kitcher P., «Ethics and Evolution. How to get here from there» In *Primates and Philosophers: How Morality Evolved*, edited by S. Macedo and J. Ober, Princeton and Oxford, Princeton University Press, 2009, p. 120-139.

- Knies, J., "Bentham on animal welfare", *British Journal for the History of Philosophy*, vol.27, n°3, 2019, pp.556-572,
- Knights, B., *The Idea of the Clerisy in the Nineteenth Century*, Cambridge: Cambridge University Press, 2010.
- Laforge R., *Psychopathologie de l'échec*, Paris : Payot, 1939.
- Laval, C., *Jeremy Bentham, les artifices du capitalisme*, Paris: Presses Universitaires de France, 2003.
- Leca-Tsiomis M., *Ecrire l'Encyclopédie. Diderot : de l'usage des dictionnaires à la grammaire*, Oxford : Voltaire Foundation, 1999.
- Lemaître J., *Jean-Jacques Rousseau*, New York : McClure Co., 1907.
- Levy, P., *Moore: G.E. Moore and the Cambridge Apostles*, New York: Holt, Rinehart and Winston, 1979.
- Lyons, D., *Rights, Welfare, and Mill's Moral Theory*, New York: Oxford University Press, 1994.
- Maas, H., *William Stanley Jevons and the Making of Modern Economics*, Cambridge: Cambridge University Press, 2005.
- Macaskill, W., *Doing Good Better: Effective Altruism and a Radical Way to Make a Difference*, London: Guardian Faber, 2015.
- Mackie, J. L., "Sidgwick's Pessimism", *The Philosophical Quarterly*, 26 (105), 1976, pp.317-327.
- Mackie J.L, *Ethics: Inventing Right and Wrong*, London, Penguin Books Limited, 1977.
- Martin, R., "A Defence of Mill's Qualitative Hedonism", *Philosophy*, 47 (180), 1972, pp.140-151.
- Mathiot J., « Politique et économie chez Jean-Jacques Rousseau », dans J. Boulad-Ayoub, I. Schulte-Jenckhoff et P-M. Vernes (dir.), *Rousseau. Anticipateur-retardataire*, Paris : L'Harmattan, 2000, pp. 19-39.

Medlin D., « André Morellet, translator of liberal thought », *Studies on Voltaire*, 1978, CLXXIV, pp.189-201.

Medlin D., « Catalogue of Morellet's works », in Medlin D., Merrick J., *André Morellet (1727-1819) in the Republic Letters and the French Revolution*, New-York : Peter Lang, 1995, pp.185-232.

Medlin D., David J.C., et Leclerc P.A., *Lettres d'André Morellet*, Oxford : Voltaire Foundation, vol. I, 1991.

Mill J.S., *Essays on Politics and Society Part I (On Liberty)*, in J.M. Robson (ed), *The Collected Works of John Stuart Mill*, Toronto: University of Toronto Press, London: Routledge and Kegan Paul, Volume XVIII, 1977.

Mill, J., and T. Ball., *James Mill: Political Writings*, Cambridge: Cambridge University Press, 1992.

Mill, J.S., *Autobiography of John Stuart Mill*, London: Floating Press, 2009.

Miller, R. B., "Actual Rule Utilitarianism", *The Journal of Philosophy*, Vol. 106, No. 1, 2009, pp. 5-28.

Mongin, P., "Harsanyi's Aggregation Theorem: Multi-Profile Version and Unsettled Questions", *Social Choice and Welfare*, 11, 1994, pp.331–354.

Mongin, P. et Cozic, M., "Rethinking nudge: not one but three concepts", *Behavioural Public Policy*, 2(1), 2018, pp. 107-124.

Moore, G.E, *Ethics*, London: Williams and Norgate, 1912.

Moore, G.E., *Principia Ethica. Revised Edition*, Cambridge: Cambridge University Press, 1993.

Morellet A., *Mémoires de l'abbé Morellet de l'Académie française sur le dix-huitième siècle et sur la Révolution*, éditées par J.P. Guicciardi, Paris : Mercure de France, 1988.

Morellet A., *Prospectus d'un nouveau dictionnaire de commerce*, Paris : Frères Estienne.

Mulgan, T., "Answering to future people", *Journal of Applied Philosophy*, 35 (3), 2018, pp.532-48.

Mulgan T., *Future People. A Moderate Consequentialism Account of our Obligations to Future Generations*, Oxford: Clarendon Press, 2006.

Mulgan, T., *The Demands of Consequentialism*, Clarendon Press, Oxford, 2001.

Murphy A., « Le développement des idées économiques en France (1750-1756) », *Revue d'Histoire Moderne et Contemporaine*, T.XXXIII, octobre-décembre, 1986, pp.521-541.

Nagel T., *The View from Nowhere*, Oxford and New York: Oxford University Press, 1989.

Nagel, T., "Rawls on Justice", *The Philosophical Review*, 82 (2), 1973, pp.220-234.

O'Donoghue T. and Rabin M., 'Doing It NOW or Later', *The American Economic Review*, 89(1), 1999, pp.103-24.

Parfit D., "Experiences, subjects, and conceptual schemes." , *Philosophical Topics*, 26 (1-1), 1999, pp.217-70.

Parfit D., *On What Matters*, Oxford: Oxford University Press, 2 vols., 2011.

Parfit, D., *On What Matters*, Oxford: Oxford University Press, vol.3, 2017.

Parfit D. , «Personal Identity and Rationality», *Synthese*, vol. 53, n°2, 1982, p.227-241

Parfit D., *Reasons and Persons*, Oxford: Clarendon Press, 1984.

Perrot J.C., « Les dictionnaires de commerce au XVIIIe siècle », *Revue d'Histoire moderne et contemporaine*, 1, 1981, pp.36-67, republié dans J.C. Perrot, *Une histoire intellectuelle de l'économie politique*, Paris : Editions de l'École des Hautes Etudes en Sciences Sociales, 1992.

Perrot J.C., *Une histoire intellectuelle de l'économie politique, 17e-18e siècles*, Paris : Éditions de l'École des hautes études en sciences sociales, 1992.

Pettit P., *Republicanism: A Theory of Freedom and Government*, Oxford : Clarendon Press, 1997.

Piguet M.F., « OEconomie/Économie (politique) dans le texte informatisé de l'Encyclopédie », *Recherches sur Diderot et sur l'Encyclopédie*, p. 31-32, Avril 2002, pp. 123-137.

Pocock J.G.A., *The Machiavellian Moment: Florentine Political Thought and the Atlantic Republican Tradition*, Princeton: Princeton University Press, 1975.

- Popper K., *The Open Society and its Enemies*, London: Routledge, 1945.
- Proust J., *Diderot et l'Encyclopédie*, Paris: Albin Michel, 1995.
- Railton, P., "Alienation, consequentialism, and the demands of morality", *Philosophy and Public Affairs*, vol.13, n°2, 1984, pp.134-171.
- Rashdall, H., *The Theory of Good and Evil*, Oxford: Clarendon Press, 1907.
- Rawls J., *A Theory of Justice*, Cambridge, Massachusetts: Harvard University Press, 1971.
- Rawls, J., *Théorie de la justice*, Traduit par C. Audard, Paris: Editions du Seuil, 1987 [1971].
- Regan, D., "How to Be a Moorean", *Ethics*, 113 (3), 2003, pp.651-677.
- Regan, T., *Bloomsbury's Prophet: G. E. Moore and the Development of His Moral Philosophy*, Eugene, Oregon: Wipf & Stock Publishers, 2012.
- Riley, J., "Interpreting Mill's Qualitative Hedonism", *The Philosophical Quarterly*, 53 (212), 2003, pp.410-418.
- Riley, J., "Is Qualitative Hedonism Incoherent?", *Utilitas*, 11 (03), 1999, pp.347-358.
- Riley, J., "On Quantities and Qualities of Pleasure", *Utilitas*, 5 (02), 1993, pp.291-300.
- Risse, M., "Harsanyi's 'Utilitarian Theorem' and Utilitarianism", *Noûs*, 36 (4), 2002, pp.550-577.
- Roemer, J. E., "Harsanyi's impartial observer is not a utilitarian", In *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*, edited by Fleurbaey M., Salles M. and Weymark J., Cambridge: Cambridge University Press, 2008, pp.129-135.
- Ross, W.D., *The Right and the Good*, Oxford: Oxford University Press, 1930.
- Rousseau J.J., *Œuvres complètes*, tome III : *Du Contrat social - Écrits politiques*, Paris : Gallimard, 1964.
- Ryan, A., *J. S. Mill*, London: Routledge and Kegan Paul, 1974.
- Ryberg J. and Tännsjö T. (eds.), *The Repugnant Conclusion. Essays on Population Ethics*, Dordrecht, Boston and London: Kluwer Academic Publishers, 2004.

Salvat C., « Histoire de la traduction inédite de la Richesse des nations par l'abbé Morellet. 'Une traduction manuscrite toujours célébrée et toujours obstinément refusée au public.' », *Storia del Pensiero Economico*, Florence, 1999, n.38, pp.119-136.

Salvat C., *Formation et diffusion de la pensée économique libérale française : André Morellet et l'économie politique du dix-huitième siècle*, Thèse de doctorat, Université Lyon 2, 2000, 2 vols.

Salvat C., “ André Morellet et la diffusion de la pensée économique italienne ”, *Il pensiero economico italiano*, Pise, VIII, 2000, n°2, pp.99-124.

Salvat C., “ Morellet André (1727-1819) ”, in Frobert L., Potier J.P., Tiran A. (sous la direction de), *Economistes en Lyonnais, en Dauphiné et en Forez*, Lyon : Institut des Sciences de l'Homme, 2000, pp.228-240.

Salvat C., van den Berg R., “La Scottish Subtlety. The commentaries of André Morellet on the Wealth of Nations ”, *European Journal of the History of Economic Thought*, 2001, vol.8, n°2, pp.146-185.

Salvat C., “L'arithmétique politique d'André Morellet ”, (en collaboration avec Pierre Crépel), in Thierry Martin (dir.), *L'arithmétique politique française*, Paris : INED, 2003, pp.195-229.

Salvat C., “Bonheur public et méthode géométrique par Marco Bianchini”, *Revue de Philosophie économique*, Juin 2003, pp.143-150.

Salvat C., “ 'De la science de l'économie publique' de Morellet ”, *Studies on Voltaire and the Eighteenth Century*, October 2003, pp.65-96.

Salvat C., “ Histoire du syntagme et du concept de division du travail au XVIIIème siècle ”, *Dictionnaire des usages socio-politiques (1770-1815)*, Paris : Champion, 2003, pp.39-66.

Salvat C., « L'affaire de la Compagnie des Indes », *Revue européenne* 2000, Décembre 2003, pp.5-6.

Salvat C., “Reason and Sentiments: Review of Emma Rothschild's Economic Sentiments. Adam Smith, Condorcet and the Enlightenment” (review article en collaboration avec Pascal Bridel), *European Journal of the History of Economic Thought*, vol.1, n°1, 2004, pp.131-145.

Salvat C., « 'On ne fait point d'industrie entre ciel et terre' : Walras, lecteur de Quesnay », *Cahiers du CERAS*, Hors série n°3, Mai 2004, pp.207-221.

Salvat C., "L'ambivalence de la référence physiocratique chez Walras", in R. Baranzini, A. Diemer et C. Mouchot (éds.), *Études walrassiennes.*, Paris: L'Harmattan, 2004, pp. 29-50.

Salvat C., "Eighteenth-Century Economics: Turgot, Beccaria and Smith and Their Contemporaries by Peter Groenewegen", *The Adam Smith Review*, volume 2, 2005, pp.275-279.

Salvat C., "Oeconomies. Les articles OE/Economie et leurs désignants dans l'Encyclopédie", *Recherches sur Diderot et l'Encyclopédie*, n°40, Octobre 2006, pp.97-116.

Salvat C., "Autonomie morale et autorité, Ou la question de la volonté chez Rousseau", *Cahiers d'Economie Politique*, n°53, 2007, 75-93.

Salvat C., "La loi et l'échange: le statut de la rationalité économique chez Rousseau", *Revue économique*, volume 58, n°2, 2007, pp.381-398.

Salvat C., "Metamorphoses of Citizenship", *History of European Ideas*, 2011, vol.37, n°2, pp.209-210.

Salvat C., "Mill et le paternalisme", Compte-rendu de Gregory Claeys, *Mill and Paternalism*, Cambridge University Press, 2013, et de David O. Brink, *Mill's Progressive Principles*, Oxford University Press, 2013, *Economies et Sociétés*, série PE, numéro 51, 2014, pp.2045-2049.

Salvat C., "Behavioral Paternalism", *Revue de Philosophie Economique*, vol.15, n°2, 2014, pp.109-130.

Salvat C., "Mill et le paternalisme", Compte-rendu de Gregory Claeys, *Mill and Paternalism*, Cambridge University Press, 2013, et de David O. Brink, *Mill's Progressive Principles*, Oxford University Press, 2013, *Economies et Sociétés*, série PE, numéro 51, 2014, pp.2045-2049.

Salvat C., "Rousseau et la « Renaissance classique » française (1898-1933) ", *Astériorion, Philosophie, Histoire des Idées, Pensée politique*, 12, 2014 [<http://asterion.revues.org/2545>]

Salvat C., "Economics of paternalism: The hidden costs of self-commanding strategies", *The Journal of Philosophical Economics. Reflections on Economic and Social Issues*, IX: 1, 2015, pp. 102-24.

Salvat C., 'Manipulation of Choice, Ethics and Libertarian Paternalism by Mark D. White', *Oeconomia, History, Methodology, Philosophy*, 5(4), 2015, pp.533-537.

Salvat C., "Parfit et les raisons du conséquentialisme de la règle", *Revue de Métaphysique et de Morale*, n°2, 2019, pp.39-50.

Salvat C., « Parfit, l'égoïsme rationnel et la question de l'identité personnelle », *Æconomia - History/Methodology/Philosophy*, 2015, n° 5-4, p. 437-460.

Salvat C., « Rétrospective des articles publiés depuis 1984 », *Economies et Sociétés*, numéro spécial 'Trente ans d'histoire de la pensée économique', 51, 2014, pp. 2015-2043.

Salvat C., 'Reasons without Persons, by B. Hedden', *Economics and Philosophy*, 32(3), 2016, pp. 548-553.

Salvat C., "French Political Economy and Positivism (or how history of economic thought became mainstream economics)", *Iberian Journal of the History of Economic Thought*, vol4., n°2, 2017. <https://revistas.ucm.es/index.php/IJHE/article/view/58093>

Salvat C., "From Literature to Politics: How Rousseau Has Come to Symbolize Totalitarianism", *Journal of Interdisciplinary History of Ideas*, 2017, vol.6, n°11, pp.4-23.

Salvat C., "Living by Default". Book review of 'Choosing Not to Choose: Understanding the Value of Choice', by Cass R. Sunstein, *Journal of Economic Methodology*, 2017, vol.24, n°4, pp.430-434.

Salvat C., "Rousseau's virtuous economics", in R. Kuroki and Y. Ando (Eds.) *The Foundations of Political Economy and Social Reform: Economy and Society in Eighteenth Century France* (Routledge Studies in the History of Economics), London: Routledge, 2018, pp.134-154.

Salvat C., "Représenter la volonté générale: Rousseau et son héritage paradoxal" [with Théophile Pénigaud de Mourgues], in M. Albertone et D. Castiglione, *Les métamorphoses de la représentation. Langages et pratiques politiques à l'époque moderne*, Paris, Classiques Garnier, 2018, pp.115-137.

Salvat C., "Egoïsme, altruisme et justice distributive", In *Dictionnaire des inégalités*, ed. P. Savidan, Paris: Presses Universitaires de France, 2018, pp.433-447.

Salvat C., "Derek Parfit", in P. Savidan (dir.), *Dictionnaire des inégalités*, Paris, Puf, 2018, pp.127-1221.



- Salvat C., "The Cosmos of duty. Henry Sidgwick's Methods of Ethics, by Roger Crisp", *Oeconomia, History, Methodology, Philosophy*, 8(4), 2018, pp.533-537.
- Salvat C., "Parfit", in *Encyclopédie Philosophique*, janvier 2019, <http://encyclophilo.fr/parfit-a-2/>.
- Salvat C., "Parfit et les raisons du conséquentialisme de la règle", *Revue de Métaphysique et de Morale*, n°2, 2019, pp.37-48.
- Salvat C., "Des personnes aux raisons : autour de Derek Parfit" (avec Laurent Jaffro), *Revue de Métaphysique et de Morale*, n°2, 2019, pp.3-6.
- Salvat C., « L'éthique face à la GPA », *Raison Publique*, n°23, 2019, pp.163-173.
- Salvat C., "Still-Born Yet not Without Influence: What Mill's Political Economy Owes to His Project of Ethology", *Journal of the History of Economic Thought*, en révision.
- Salvat C., *L'utilitarisme*, La Découverte, sous contrat, parution prévue 2020.
- Samuelson, P.A., "A Note on the Pure Theory of Consumer Behavior", *Economica*, 5 (17), 1938, pp.61-71.
- Scanlon T.M, *What We Owe to Each Other*, Cambridge, Massachusetts and London, England: The Belknap Press of Harvard University Press, 1998
- Scarre, G., "Donner and Riley on Qualitative Hedonism", *Utilitas*, 9 (03), 1997, pp.351-360.
- Scheffler, S., *The Rejection of Consequentialism: a Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*, Oxford-New York: Clarendon Press-Oxford University Press, 1982.
- Scheffler, S., and N. Kolodny, *Death and the Afterlife*, New York: Oxford University press, 2013.
- Schelling T.C., 'Self-Command: A New Discipline', in G. Loewenstein and J. Elster (eds.), *Choice over Time*, New York: Russell Sage Foundation, 1992, pp.167-76.
- Schelling T.C., "Self-Command in Practice, in Policy, and in a Theory of Rational Choice", *The American Economic Review*, 74:2, 1984, pp.1-11.
- Schmidt, C. "Psychologie cognitive et analyse des décisions économiques", *Revue française d'économie*, 20 (3), 2006, pp.3-51.

Schneewind J. B., "Sidgwick and the Cambridge Moralists", *The Monist*, 58 (3), 1974, pp.371-404.

Schneewind J.B., 'Sidgwick and the Cambridge moralists', in B. Schultz (ed.), *Essays on Henry Sidgwick*, Cambridge: Cambridge University Press, 1992, pp. 93-121.

Schneewind, J. B. , *Sidgwick's Ethics and Victorian Moral Philosophy*, Oxford: Oxford University Press, 1977.

Schultz B., *Henry Sidgwick - Eye of the Universe: An Intellectual Biography*, Cambridge: Cambridge University Press, 2004.

Sen A., *Collective Choice and Social Welfare*, San Francisco: Holden Day, 1970

Sen, A.K., "Non-linear Social Welfare Functions: A Reply to Professor Harsanyi", In *Foundational Problems in the Special Sciences*, edited by Butts R. E. and Hintikka J, Dordrecht: Springer, 1977, pp. 297–302.

Sen, A.K., "The Impossibility of a Paretian Liberal." *Journal of Political Economy* 78 (1), 1970, pp.152–157.

Sen, A.K., "Utilitarianism and Welfarism", *Journal of Philosophy*, 76 (9), 1979, pp.463– 488.

Sen, A.K., "Equality of What?" In Tanner Lectures on Human Values, Volume 1, edited by S. McMurrin. Cambridge: Cambridge University Press, 1980.

Shackleton, R., "The Greatest Happiness of the Greatest Number: The History of Bentham's Phrase", *Studies on Voltaire and the Eighteenth Century*, 90, 1972, pp.1461-1482.

Shaw, W., *Contemporary Ethics: Taking Account of Utilitarianism*, Londres: Wiley-Blackwell, 1999.

Shaw, W., *Utilitarianism and the Ethics of War*, Londres: Routledge, 2016.

Shoemaker D., 'Parfit on Identity', in J. Dancy, *Reading Parfit*, Oxford: Blackwell Publishers, 1997, pp.135-48.

Shoemaker S., "Personal Identity and Ethics" In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta, 2014. <<http://plato.stanford.edu/archives/spr2014/entries/identity-ethics/>>.

- Shoemaker S., "Persons and Their Pasts", *American Philosophical Quarterly*, 7, 1970, pp.269–285.
- Shoemaker S. and R. Swinburne, *Personal Identity. The Dualist Theory*, Oxford: Basic Blackwell, 1984.
- Sidgwick, H., *The Methods of Ethics*, Seventh Edition ed. Indianapolis/Cambridge: Hackett Publishing Company, 1981.
- Sigot, N., *Bentham et l'économie: une histoire d'utilité*, Paris: Economica, 2001.
- Simon, H., "A Behavioral Model of Rational Choice", *Quarterly Journal of Economics*, 69 (1), 1955, pp.99-118.
- Singer P., "Famine, Affluence, and Morality", *Philosophy and Public Affairs*, 1(1), 1972, pp. 229-243.
- Singer, P., *La libération animale*, traduit par L. Rousselle, Paris : Petite Bibliothèque Payot, 2012.
- Singer, P., *Practical Ethics*, Cambridge: Cambridge University Press, 1993.
- Singer, P., *The Most Good You Can Do: How Effective Altruism Is Changing Ideas About Living Ethically*, New Haven and London: Yale University Press, 2015.
- Singer, P., *Théorie du tube de dentifrice. Comment changer le monde selon Henry Spira*, traduction par A. Pons, Paris : Editions Goutte d'Or, 2018.
- Skelton, A., "Ideal Utilitarianism: Rashdall and Moore", In *Underivative Duty: British Moral Philosophers from Sidgwick to Ewing*, edited by T. Hurka, Oxford: Oxford University Press, 2011, pp. 45–65.
- Skinner Q., *The Foundations of Modern Political Thought*, Cambridge: Cambridge University Press, 1978.
- Skorupski J., *John Stuart Mill*, Londres: Routledge & Kegan Paul, 1989.
- Skorupski, J., "Three Methods and a Dualism", *Proceedings of the British Academy*, (109), 2001, pp.61-81.
- Slote, M.A., *Goods and Virtues*, Oxford: Clarendon Press, 1983.

Slote, M.A., *Common-Sense Morality and Consequentialism*, Londres: Routledge & Kegan Paul, 1985.

Slote, M. et P. Pettit, "Satisficing Consequentialism", *Proceedings of the Aristotelian Society*, Supplementary Volumes 58, 1984, pp.139-176.

Smart, J.J.C., et B. Williams, *Utilitarianism: For and Against*, Cambridge: Cambridge University Press, 1973.

Smart, R. N., "Negative Utilitarianism", *Mind*, 67 (268), 1958, pp.542-543.

Smith V.L., "Adam Smith on Humanomic Behavior", *The Journal of Behavioral Finance & Economics*, Volume 2, Issue 1, Spring 2012, pp.1-20.

Sober E. et Wilson D., *Unto Others: The Evolution and Psychology on Unselfish Behavior*, Cambridge, Harvard University Press, 1998.

Sonenscher M., *Before the Deluge: Public Debt, Inequality, and the Intellectual Origins of the French Revolution*, Princeton: Princeton University Press, 2009

Spector C., *Au prisme de Rousseau : usages politiques contemporains*, Oxford : Voltaire Foundation, vol. VIII, 2011.

Spitz J.F., *La liberté politique: essai de généalogie conceptuelle*, Paris : Presses Universitaires de France, 1995.

Spitz, J.-F., "Le conséquentialisme hybride de Samuel Scheffler : défense et illustration", in S. Guerard De Latour et M.-A. Dilhac (eds.), *Étant donné le pluralisme*, Paris : Éditions de la Sorbonne, 2013, pp.27-53.

Starobinski J., *Jean-Jacques Rousseau: la transparence et l'obstacle ; suivi de Sept essais sur Rousseau*, Paris : Gallimard, 1989.

Strotz R.H., 'Myopia and Inconsistency in Dynamic Utility Maximization', *The Review of Economic Studies*, 23 (3), 1955, pp.165-180.

Sugden R., "Why incoherent preferences do not justify paternalism", *Constitutional Political Economy*, 19:3, pp. 226-248.

Sumner, L. W., *Welfare, Happiness, and Ethics*, Oxford: Oxford University Press, 1996.

Sumner, L. W.. *Welfare, Happiness, and Ethics*. New York-Oxford: Oxford University Press.

- Sumner, L.W., *Abortion and moral theory*, Princeton, New Jersey: Princeton University Press, 1981.
- Sunstein C., Thaler R., "Libertarian Paternalism Is Not an Oxymoron", *The University of Chicago Law Review*, 70:4, 2003, pp. 1159-1202.
- Talmon, J.L., *The Origins of Totalitarian Democracy*, London: Secker&Warburg, 1955.
- Tännsjö, T., *The Relevance of Metaethics to Ethics*, Stockholm: Almqvist & Wiksell International, 1976.
- Tännsjö T., "Why We Ought to Accept the Repugnant Conclusion", *Utilitas*, 14 (3), 2002, pp.339 - 359.
- Temkin, L. S., "Intransitivity and the Mere Addition Paradox." *Philosophy and Public Affairs* 16 (2), 1987, pp.138–187.
- Thaler R., "Toward a Positive Theory of Consumer Choice", *Journal of Economic Behavior and Organization*, 1(1), 1980, pp. 39-60.
- Thaler R.H., Sunstein C.R, *Nudge. Improving Decisions About Health, Wealth, and Happiness*, New Haven & London: Yale University Press, 2008.
- Thomas, W., "James Mill's Politics: The 'Essay on Government' and the Movement for Reform", *The Historical Journal*, 12 (2), 1969, pp.249-284.
- Thomas, Wi., "L'utilitarisme et le libéralisme anglais au début du XIXe siècle." In *Regards sur Bentham et l'utilitarisme* edited by Kevin Mulligan and Robert Roth, Genève: Librairie Droz, 1993, pp.39-58.
- Toulmin, S.E., *An Examination of the Place of Reason in Ethics*, Cambridge: Cambridge University Press, 1950.
- Trousseau R., *Défenseurs et adversaires de Jean-Jacques Rousseau. D'Isabelle de Charrière à Charles Maurras*, Paris : Honoré Champion, 1995.
- Tversky A., Kahneman D., "The Framing of Decisions and the Psychology of Choice", *Science*, 211: 4481, 1981, pp. 453-458.
- Tversky A., Slovic P., Kahneman D., "The Causes of Preference Reversal", *The American Economic Review*, 80:1, 1990, pp. 204-17.

Urmson, J. O., "The Interpretation of the Moral Philosophy of J. S. Mill", *The Philosophical Quarterly*, vol. 3, n°10, 1953, pp.33-39.

Villey D., « Prolégomènes à l'enseignement de la philosophie économique », *Revue d'Economie Politique*, Mai-Juin 1959, pp. 362-414.

Viroli M., *Jean-Jacques Rousseau and the 'Well-Ordered Society'*, Cambridge: Cambridge University Press, 1989,

Weber H., *La Compagnie française des Indes, 1604-1875*, Paris : Arthur Rousseau Editeur, 1904.

West, H.R., "Mill's Qualitative Hedonism", *Philosophy*, 51 (195), 1976, pp.97-101.

West, H.R., *An Introduction to Mill's Utilitarian Ethics*, Cambridge: Cambridge University Press, 2004.

Weymark, J. A., "A Reconsideration of the Harsanyi–Sen Debate on Utilitarianism" In *Interpersonal Comparisons of Well-being*, edited by Elster J. and Roemer J. E, Cambridge: Cambridge University Press, 1991, pp. 255–320.

Wiggins D., *Identity and spatio-temporal continuity*, London: Blackwell, 1967.

Williams, B., "A Critique of Utilitarianism", in Smart J.J.C et Williams B., *Utilitarianism: For and Against*, Cambridge: Cambridge University Press, 1973, pp.75-150.

Williams B., *Problems of the Self: Philosophical Papers 1956-1972*, Cambridge: Cambridge University Press, 1976.

Winch D., *Riches and Poverty: An Intellectual History of Political Economy in Britain, 1750-1834*, Cambridge: Cambridge University Press, 1996.



# Sélection de travaux (I)

## *L'utilitarisme*

Ouvrage en cours de publication, La découverte, collection Repères, automne 2020.





# L'Utilitarisme

Christophe Salvat

Collection 'Repères'

La Découverte

*Pour Alex,*

Remerciements : Je souhaite remercier le rapporteur anonyme de La Découverte pour ses nombreuses remarques et suggestions. Je remercie également Jean Baccelli, Cyril Hédoïn, Sylvie Rivot, et Antonin Thil pour leur relecture. Merci enfin à tous mes étudiants, à Cambridge, Lyon et Aix-en-Provence pour leurs questions et remarques au cours de ces dernières années. Toutes les erreurs restent naturellement miennes.

# Table des Matières

## INTRODUCTION

*Qu'est-ce que l'utilitarisme ?*

*Les raisons d'un discrédit*

*Pourquoi s'intéresser à l'utilitarisme aujourd'hui ?*

*Objet et plan de l'ouvrage*

## I/ L'UTILITARISME CLASSIQUE

### **Le programme utilitariste de Bentham**

*Jeremy Bentham (1798-1832), le philosophe réformateur*

*Le principe d'utilité*

*Le calcul de félicité*

### **L'utilitarisme libéral de John Stuart Mill**

John Stuart Mill (1806-1873), figure intellectuelle du XIX<sup>e</sup> siècle

*Encadré 1 : John Stuart Mill et Harriet Taylor*

L'hédonisme qualitatif de Mill

L'utilitarisme libéral de Mill

### **Les Méthodes d'Henry Sidgwick**

Henry Sidgwick (1838-1900), le professeur de Cambridge

Les trois méthodes de l'éthique

Le dualisme de la raison pratique

## II/ MESURER, AGGREGER ET MAXIMISER L'UTILITE

### **Les trois conceptions du bien-être**

La théorie des états mentaux

A théorie du succès et l'utilitarisme des préférences

La théorie des listes objectives

### **L'utilité comme mesure du bien**

Théories cardinales et ordinales de l'utilité

*Encadré 2. Mesure de l'utilité et théories du bien-être*

Utilité totale ou utilité moyenne ?

### **Questions relatives à la maximisation d'utilité**

Sur quel horizon temporel maximiser l'utilité ?

L'utilitarisme négatif

Doit-on absolument maximiser l'utilité ?

## **III/ L'UTILITARISME DE LA REGLE**

### **De l'utilitarisme de l'acte à l'utilitarisme de la règle**

Qu'est-ce que l'utilitarisme de la règle ?

Encadré 3. Utilitarisme de la règle et généralisation utilitariste

### **Principaux modèles d'utilitarisme de la règle**

L'utilitarisme à deux niveaux de Hare

Le code moral idéal

Utilitarisme de la règle et contractualisme

### **Avantages et limites de l'utilitarisme de la règle**

L'utilitarisme de la règle face aux objections faites à l'utilitarisme classique

Limites de l'utilitarisme de la règle

Au-delà de l'utilitarisme de la règle

## **IV/ IMPARTIALITE : THEORIE ET PRATIQUE**

### **Le principe d'impartialité**

Impartialité, universalité et égalité

Impartialité et maximisation d'utilité

Impartialité et égalité

### **Limites du principe d'impartialité**

De l'impartialité à l'impersonnalisme

Prendre les droits des personnes au sérieux

L'utilitarisme est-il vraiment égalitaire ?

Encadré 4. Le système des enchères de Dworkin

### **(Ré)Intégrer la notion de personne**

(Ré)concilier utilité et droits

Le prioritarisme

Existe-t-il des prérogatives personnelles en morale ?

### **L'impartialité en pratique : l'utilitarisme face à l'éthique appliquée**

Les famines et la grande pauvreté

L'éthique animale

L'éthique environnementale

## **CONCLUSION**

## **REPERES BIBLIOGRAPHIQUES**

## INTRODUCTION

*Qu'est-ce que l'utilitarisme ?*

L'utilitarisme est une philosophie morale née au dix-huitième siècle sous l'impulsion de Jeremy Bentham (1748-1832), qui souhaitait donner des fondements rationnels à la morale et à la justice. Bentham entendait ainsi s'attaquer à la fois à la tradition, à la jurisprudence et à l'intuitionnisme comme principes d'évaluation morale, qu'il considérait arbitraires et iniques.

La philosophie utilitariste classique repose sur quatre principes fondateurs : le conséquentialisme, selon lequel une action est moralement bonne (*right*) si ses conséquences sont avantageuses pour l'ensemble de la communauté, l'hédonisme, selon lequel une action est considérée comme avantageuse si elle augmente le plaisir (net des peines) des individus, l'impartialité, selon laquelle tous les intérêts des individus doivent être traités également, et le maximalisme, soit l'idée qu'une fois le critère d'évaluation morale connu, les agents sont tenus d'agir de façon à en maximiser la quantité. Ces principes ne sont pas, cependant, toujours caractéristiques de l'utilitarisme contemporain. Enfin, précisons que dans sa version traditionnelle, la théorie utilitariste est une procédure d'évaluation *directe* des actions *individuelles*. Dans un certain nombre de ses versions contemporaines, elle se présente comme une théorie d'évaluation *indirecte* des actions *individuelles* ou comme une théorie d'évaluation *directe* des actions *collectives*.

La pensée utilitariste est, dans l'esprit de son fondateur, une philosophie globale destinée à s'appliquer à toute une variété de domaines de la vie publique, que ce soit la constitution politique, l'organisation institutionnelle, l'économie politique, l'enseignement scolaire etc., ou de la vie privée, l'éducation des enfants, la religion, le mariage, etc. Aujourd'hui, même si l'utilitarisme reste un cadre d'analyse de référence en éthique appliquée, sa dimension

juridique, politique et institutionnelle est quasi- absente et rarement assumée. Le champ de l'utilitarisme se limite donc essentiellement à celui de la philosophie morale.

Contrairement à l'image que l'on en donne souvent, la philosophie utilitariste a joué depuis son origine défendu des valeurs généralement considérées comme progressistes, que ce soit la défense du système de démocratie représentative et l'élargissement du droit de vote aux femmes, notamment défendu par J.S. Mill, ou une plus juste répartition des richesses, plus propice à la maximisation de l'utilité totale (du fait du principe de décroissance de l'utilité marginale), ou encore une plus grande tolérance envers les minorités (les femmes, les pauvres, les personnes racisées, les homosexuels etc.). Les philosophes utilitaristes, le plus souvent d'origine anglo-saxonne, ont également été d'importants représentants de la pensée libérale, à l'instar de John Stuart Mill. Mais l'utilitarisme et le libéralisme ne vont pas nécessairement de pair. La maximisation de l'utilité totale requiert souvent une politique fortement interventionniste de la part du gouvernement, ce qui est difficilement réconciliable avec le principe de non-nuisance (*harm principle*), développé par Mill dans *On Liberty*, qui est la règle d'or du libéralisme. L'articulation entre ces deux orientations politiques n'est donc pas toujours facile.

#### *Les raisons d'un discrédit*

Dans sa postface au premier volume de la réédition de *La formation du radicalisme philosophique* d'Elie Halévy, initialement parue entre 1901 et 1904, et republiée en 1995, Jean-Pierre Dupuy relève le peu d'intérêt que suscite la philosophie utilitariste en France : « « Que savez-vous de la doctrine utilitaire ? » demande à ses compatriotes le Professeur Halévy dès les premières lignes de son ouvrage. La même question, posée quelque quatre-vingt-dix ans plus tard, donnerait des notes encore plus désastreuses que celles qu'il leur décerne en 1901. De cette doctrine, qui, pourtant, exerça une influence sur la Révolution française, si l'on veut bien compter Helvétius parmi ses précurseurs, nous ne savons rien, parce que nous ne voulons rien en savoir. » [Dupuy, 1995, p.332]

Bien que l'œuvre de Bentham ait d'abord reçu un accueil favorable en France, dans la traduction et l'important travail d'édition réalisé par Etienne Dumont [Blamires, 2008 ; de



Champs, 2015 ; Whatmore 2007] avant de se diffuser en Angleterre, la philosophie utilitariste ne s'est jamais véritablement imposée et est souvent considérée avec méfiance. Pour quelles raisons l'utilitarisme a-t-il aujourd'hui aussi mauvaise réputation ? Comme souvent celle-ci repose d'abord sur des préjugés nourris par la méconnaissance que l'on a de cette philosophie. Mais cela n'explique pas tout. L'utilitarisme a été l'objet, dès ses origines, de virulentes attaques de la part de ses opposants, des représentants religieux bien-sûrs, aux écrivains romantiques en passant par les défenseurs de l'intuitionnisme moral ou du conservatisme politique. En France, pays dans lequel l'héritage des Lumières aurait pu faire présumer d'un meilleur accueil, l'utilitarisme a beaucoup souffert de la domination sans partage du positivisme d'Auguste Comte.

L'utilitarisme y a toujours mauvaise presse et est régulièrement associé avec l'économisme, l'égoïsme et le matérialisme. Il est vrai que les théories économiques du XX<sup>e</sup> siècle se sont beaucoup inspirées du cadre théorique fourni par l'utilitarisme pour défendre une vision 'économiste' du bien-être social. Cela a d'ailleurs conduit à la création en 1981 d'un Mouvement Anti-Utilitariste en Sciences Sociales (MAUSS), ainsi qu'à celle de la *Revue du MAUSS* (semestrielle, La Découverte) et du *Journal du MAUSS* (en ligne). A cela s'ajoute l'affaiblissement général de l'utilitarisme face au renouveau libéral des années 1970 et 80 aux Etats-Unis, mené notamment par Rawls, Nozick et Dworkin.

Enfin, l'enseignement des théories utilitaristes dans les facultés de philosophie, de droit ou d'économie se réduit souvent à une présentation de ses fondateurs, Bentham et Mill, qui sont à ce jour les seuls auteurs utilitaristes traduits en français. Pourtant la philosophie utilitariste s'est considérablement développée et perfectionnée depuis le XIX<sup>e</sup> siècle. L'utilitarisme n'est pas une doctrine du passé. Les débats suscités par l'éthique appliquée et la méta-éthique, sans oublier naturellement les objections des théories rivales de la philosophie utilitariste (déontologie, contractualisme et éthique de la vertu), ont considérablement fait évoluer cette dernière. La philosophie utilitariste a, on le verra, beaucoup évolué depuis le XVIII<sup>e</sup> siècle et elle évolue encore, au gré des objections théoriques qui lui sont adressées, mais également des enjeux sociaux éthiques auxquels les philosophes et les politiques sont confrontés.

Le développement de l'éthique appliquée dans les années 1970, sous l'impulsion de Peter Singer et de Jonathan Glover notamment, a beaucoup contribué à son renouvellement. En tant que philosophie morale, l'utilitarisme se doit de prendre part aux débats moraux qui secouent notre société. Dans les années 1970 et 80, les philosophes ont été appelés à se prononcer sur la légitimité morale de l'avortement, de l'euthanasie ou de la guerre. Aujourd'hui, ils sont consultés sur les politiques à adopter face au réchauffement climatique, la souffrance ou l'exploitation animale ou encore la procréation assistée (GPA-PMA). Certaines de ces questions exigent parfois d'élargir ou de modifier le cadre théorique de référence, par exemple en incluant/renforçant la dimension intergénérationnelle et/ou antispéciste dans le modèle de base. Mais l'évolution de la théorie normative peut également être le produit de débats ayant lieu en amont, au niveau de la théorie méta-éthique [Tännsjö, 1976]. On sous-estime trop souvent le rôle joué par certains philosophes tels G.E. Moore, qui, au début du XX<sup>e</sup> siècle, dénonça le sophisme naturaliste des utilitaristes 'classiques', qui cherchaient à caractériser le 'bien' par les propriétés naturelles des objets désirés [Moore 1903, 1912 ; O'Connor 1982]. Or, le propre du bien moral, selon Moore, est de ne pouvoir être réduit à des propriétés non morales (comme le plaisir). Les écrits de Moore ont eu un impact considérable sur les théoriciens de l'utilitarisme du XX<sup>e</sup> siècle, tel que Richard Hare, mais également sur des philosophes contemporains tels que Peter Singer, Derek Parfit ou Jonathan Dancy [Dancy, 2006 ; Lazari-Derek and Singer, 2014 ; Parfit, 2011].

### *Pourquoi s'intéresser à l'utilitarisme aujourd'hui ?*

Les questions éthiques qui se posent à nous aujourd'hui sont nombreuses. L'évolution technologique, de plus en plus rapide, associée à la mutation des normes sociales, soulève un certain nombre de controverses morales [Shaw, 1999]. Parmi les plus actuelles, citons la bioéthique [Glover, 2017 ; Brock, 1993 ; Singer, 1993], l'éthique climatique [Broome, 2012], l'éthique des robots [Anderson et Anderson, 2011], l'éthique de la guerre [Shaw, 2016], l'éthique des affaires [Gustafson, 2013] ou encore l'éthique de la population [Arrhenius, Ryberg et Tännsjö, 2017]. Bien que prenant des formes nouvelles, un grand nombre de ces débats se réduisent *in fine* à des débats éthiques plus anciens portant sur

les fins de la médecine et de la technologie, la perfectibilité de l'homme ou encore le fondement du naturalisme moral. On pourrait donc arguer que les théories éthiques existantes suffisent à nourrir, sinon trancher, les débats. Il existe, cependant, au moins une grande question éthique qu'aucune grande philosophie morale n'a su anticiper, celle du réchauffement climatique et de ses conséquences probables sur la biodiversité et le futur de l'espèce humaine. La possibilité, à plus à moins long terme de l'extinction de l'espèce humaine, ainsi que de nombreuses autres espèces animales, et la certitude, qu'à court et moyen terme, les générations futures auront à subir les conséquences dramatiques du dérèglement climatique ainsi que de la dégradation des écosystèmes, constitue sans aucun doute une question éthique de nouvel ordre, et peut-être la plus fondamentale d'entre toutes.

### *Objet et plan de l'ouvrage*

L'utilitarisme est un ensemble théorique d'une extrême diversité et d'une grande richesse, et qui au fil du des années, s'est considérablement inspiré de ses théories rivales, la déontologie kantienne, le contractualisme et, dans une moindre mesure, l'éthique de la vertu. Ce livre présente l'utilitarisme, notamment dans sa dimension philosophique, avec une attention appuyée sur les débats traversants la philosophie morale contemporaine, afin de fournir aux lecteurs des outils conceptuels pour bien analyser les questions éthiques qui se posent à eux.

L'ouvrage est structuré autour des principales évolutions de l'utilitarisme. Il débute par une présentation de l'utilitarisme classique que l'on associe généralement à la pensée de Bentham, Mill et Sidgwick (chapitre 1). Il introduit ensuite (chapitre 2) les questionnements relatifs à la théorie de la valeur utilisée (hédonisme, théorie des préférences, liste objective), ainsi qu'au principe de maximisation (maximisation totale ou moyenne de l'utilité, utilitarisme modéré, utilitarisme négatif, prioritarisme). Il présente alors le passage de l'utilitarisme de l'acte à l'utilitarisme de la règle, censé intégrer une dimension collective et/ou déontologique dans le raisonnement moral (chapitre 3). Le dernier chapitre est consacré au principe d'impartialité, à ses implications théoriques et à sa remise en cause,

ainsi qu'à ses implications pratiques en éthique appliquée, à travers trois exemples, celui des famines, celui de l'éthique animale et celui de l'éthique environnementale.

# I/ L'UTILITARISME CLASSIQUE

On associe généralement la pensée utilitariste classique à celle de ses fondateurs, et principalement celle de Jeremy Bentham (1748-1832). Nous aurons l'occasion de voir que la théorie utilitariste contemporaine ne ressemble que très peu à celle imaginée par Bentham, il y a maintenant plus de deux siècles. Néanmoins, une bonne connaissance des sources historiques reste essentielle pour comprendre les fondements théoriques et mesurer l'ampleur de l'évolution de cette philosophie. La vie et l'œuvre des principaux fondateurs de l'utilitarisme – Bentham, Mill et Sidgwick – ont fait l'objet de très nombreuses publications (voir, par exemple, [Albee, 1902 ; Crimmins, 2014 ; Crisp 2014, 2015 ; Harrison 1983 ; Rosen, 2003 ; Driver, 2014b ; Ryan, 1974 ; Schofield, 2006, 2009 ; Schultz, 2017 ; Skorupski, 1989 ; West, 2007, 2014]), y compris en français [Audard, 1999 ; Halévy, 1901-1904 ; Bozzo-Rey et Dardenne 2011 ; Sigot 2011].

## **Bentham et le radicalisme philosophique**

*Jeremy Bentham (1748-1832), le philosophe réformateur*

Jeremy Bentham est unanimement reconnu comme le père de l'utilitarisme. C'est, en effet, lui qui en a posé les principes et qui a œuvré toute sa vie pour sa reconnaissance, sa diffusion, son application et sa postérité. La renommée de Bentham est assez tardive. Il est, tout d'abord, essentiellement connu pour ses critiques de Blackstone (*Fragment on Government*, 1776) et de Smith (*Defence of Usury*, 1787), qui toutes deux n'ont rencontré qu'une réception mitigée. Son *Introduction to the Principles of Morals and Legislation*, imprimé en 1780, mais seulement publié en 1789, ne fut pas même commentée dans les journaux. Il doit ses premiers succès éditoriaux à l'intervention du Genevois (et ancien rousseauiste) Etienne Dumont, qui traduisit (et édita) une partie de son œuvre [Blamires, 2008 ; de Champs, 2015 ; De Champs et Cléro 2009 ; Whatmore 2007].

La doctrine de l'utilité compte parmi ses précurseurs un certain nombre de penseurs européens du XVII<sup>e</sup> et du XVIII<sup>e</sup> siècle. D'une certaine manière, elle est le produit de la philosophie morale moderne [Caillé, Lazzeri et Senellart, 2001]. Bentham reconnaît notamment l'influence du philosophe matérialiste français Claude-Adrien Helvétius (1715-1771), dont il compare l'influence dans les sciences morales à celle de Bacon dans les sciences physiques, et du philosophe et juriste italien Cesare Beccaria (1738-1794), dont l'ouvrage *Des Délits et des Peines* (1764) est aujourd'hui considéré comme le véritable fondement du droit pénal moderne. La place de David Hume dans l'histoire de l'utilitarisme a soulevé beaucoup de discussions [Albee, 1902 ; Cléro, 2013 ; Halévy, 1901 ; Rosen 2003]. D'un côté, son approche méthodologique des sciences de l'homme, inspirée par la méthode expérimentale des sciences de la nature, et de l'autre, sa théorie des perceptions, en font un précurseur idéal à la théorie de l'utilité. Son scepticisme moral et sa critique du rationalisme moral, en revanche, sont difficilement compatibles avec le projet utilitariste.

En dépit de la forte dimension théorique de son œuvre, Bentham veut d'abord être un réformateur [De Champs, 2008 ; Postema, 2019 ; Schofield, 2006 ; Zhai et Quinn, 2014]. Il s'attaque notamment aux lois constitutionnelles, civiles et pénales anglaises qu'il juge injustes et inadaptées. Dans son *Fragment on Government* (1776), il condamne le système du *Common Law*, et son principal défenseur, William Blackstone, comme étant un système inique produit par une classe de juristes dans le seul but d'asseoir leur autorité et de conserver leurs prérogatives. C'est donc une réforme juridique globale et profonde que Bentham appelle de ses vœux. Il souhaite lui substituer une justice démocratique, égalitaire et fondée sur le seul principe de l'utilité.

Bentham établit un parallèle entre la loi et la morale. L'une et l'autre ne peuvent être justifiées ni par la tradition ni par la fiction. La tradition, tout d'abord, n'est pas toujours bonne ni toujours juste, et il serait absurde de vouloir conserver une mauvaise loi simplement parce que celle-ci est ancienne. Le recours à la tradition en matière pénale, ensuite, exige la comparaison de faits nécessairement dissimilaires. Elle donne un pouvoir discrétionnaire au juge et rend la sentence incertaine. Or, selon Bentham, chaque justiciable doit pouvoir anticiper avec précision la sanction auquel il s'expose pour correctement se décider. Les lois pénales doivent être aussi simples et claires que possible si l'on veut qu'elles soient appliquées. L'usage de la fiction en matière juridique participe

également pour Bentham de l'obscurité et, par conséquent, de l'inefficacité des lois. Les fictions juridiques, particulièrement utilisées dans le système de droit coutumier anglais (*Common Law*), sont des artifices techniques destinés à assimiler des cas *a priori* différents et donc à étendre le champ d'une loi ancienne à des cas nouveaux. En plus d'être source de grande incertitude pour les justiciables, la fiction juridique donne un pouvoir exagéré aux juges, dont la fonction est d'appliquer et non de créer la loi au grès de leurs interprétations [Guillot, 2014 ; Laval, 1994 ; Tusseau 2014].

### *Le principe d'utilité*

Le principe d'utilité, auquel Bentham préférera plus tard l'appellation 'principe du plus grand bonheur' [Cléro, 2006, p.23-36], se présente comme une alternative aux systèmes moraux traditionnels, celui de l'ascèse stoïcienne et religieuse, d'une part, et celui de la sympathie d'autre part. Bentham voit dans le principe de l'ascétisme un principe directement opposé à celui de l'utilité. Alors que le principe de l'utilité a pour objet le plus grand bonheur des individus, celui de l'ascèse – selon Bentham – approuve les actions dans la mesure où elles diminuent leur bonheur. Quant au principe de sympathie, Bentham lui reproche d'être arbitraire car, à défaut de critère objectif d'évaluation, il repose sur la seule approbation ou désapprobation de l'opinion publique. Bien qu'encore en vogue à la fin du dix-huitième siècle, Bentham considère la doctrine de la sympathie comme un vestige du passé, dont les fondements ont été désavoués par les philosophes des Lumières, et auxquels il se propose de substituer les méthodes rationnelles de la doctrine de l'utilité. Il rêve de devenir le Newton du monde moral [Cot, 1992].

Le principe d'utilité répond aux exigences d'une loi juste et efficace. Son objet, la maximisation du bonheur total, est rationnel et égalitaire. Il ne sert aucune classe ni aucun individu en particulier. Comme l'écrira Peter Singer, « un intérêt est un intérêt quel que soit la personne dont il est l'intérêt » [Singer, 1997, p.32]. Cela ne signifie pas que tous les individus doivent être traités de la même manière. La règle de l'impartialité, formulée par Bentham dans *Rationale of Judicial Evidence* (1825), impose parfois une différenciation de traitement. C'est le cas, notamment, en matière pénale, où le principe utilitariste exige que

les peines soient adaptées aux circonstances et aux sensibilités des individus [Guillot, 2011]. Nous verrons dans le chapitre 4 que le principe d'impartialité peut, dans certains cas, déroger à celui de l'égalité.

La règle de l'impartialité implique pour le législateur de suivre une méthode rigoureuse. Le législateur, écrit Bentham dans sa *Déontologie*, « doit être impassible comme le géomètre. Tous deux résolvent des problèmes à l'aide de calculs dépassionnés. » [Bentham, 1834, vol.2, p.26]. Ne pas s'autoriser à être sensible est, pour le législateur, une condition nécessaire pour apprécier les sentiments des autres, de tous les autres, à leur juste valeur. Le calcul est, pour les utilitaristes, un instrument de mesure de la sensibilité.

Le législateur est appelé à jouer un rôle primordial chez Bentham, plus que chez n'importe quel autre utilitariste. L'importance de la politique et de la justice dépend de la capacité des individus à respecter d'eux-mêmes les règles morales. Moins ceux-ci sont à-mêmes de s'auto-gouverner, plus fort est le besoin d'établir un système « d'harmonisation artificielle des intérêts individuels », selon l'expression d'Elie Halévy [Halévy, 1901-1904]. Celle-ci soulève, cependant, la délicate question de l'articulation des dimensions éthiques et législatives de la théorie de l'utilité de Bentham [Mongin, 1995 ; Tusseau, 2019].

Selon Bentham, il existe quatre leviers, ou sanctions, capables d'infléchir l'égoïsme des individus : la sanction *physique*, la sanction *morale* ou *populaire*, la sanction *religieuse* et la sanction *politique*. La sanction *physique* est naturelle ou biologique. On dira ainsi que l'indigestion est la sanction physique qui pénalise l'homme incapable de modérer son hédonisme alimentaire. La sanction *morale* ou *populaire* est sociale. Elle n'est ni concertée ni parfois même consciemment voulue. Elle se manifeste généralement par le sentiment de désapprobation ou d'ostracisme suscité par certaines actions jugées trop égoïstes. La sanction *religieuse* émane, comme son nom l'indique, de Dieu ou de ses représentants sur Terre. Elle prend généralement la forme de condamnations présentes (repentance) ou futures (enfer). Le droit relève de la sanction *politique*. La réflexion de Bentham est très détaillée puisqu'il va jusqu'à proposer une réorganisation du système pénitencier autour d'un modèle de prison 'idéale', rendu célèbre par Michel Foucault, le panoptique [Foucault, 1975]. L'ensemble de ces sanctions, et leur correcte anticipation, permet en théorie aux



individus de concilier la recherche de leur plaisir personnel avec la maximisation de l'utilité totale.

La maximisation de l'utilité *via* la loi exige, cependant, une adhésion stricte du législateur à l'éthique utilitariste. Si la loi permet aux individus, de nature égoïste, de se plier à l'intérêt général, il existe un risque que les législateurs élaborent la loi en privilégiant la satisfaction de leurs propres intérêts, qualifiés de 'funestes' (*sinister interests*) par Bentham. Après avoir brièvement envisagé la possibilité d'un despote bienveillant, Bentham se rallie à James Mill (le père de John Stuart), et considère que le système parlementaire et démocratique comme le seul capable d'amener les législateurs à concevoir des lois exclusivement orientées vers le bonheur de tous [Schofield 2006].

Le système démocratique défendu par Bentham exige un système de transparence absolue et de contrôle permanent des élus par les citoyens. Elle rappelle en cela le principe constitutif de son modèle d'architecture carcérale, le panoptique : le législateur et les citoyens doivent se surveiller mutuellement.

### *Le calcul de félicité*

Le calcul de félicité ou *Felicific Calculus* est considéré comme l'outil de base du raisonnement utilitariste bien que celui-ci ne soit développé que dans le seul chapitre 4 de *l'Introduction aux principes de morales et de législation* (1789). Il permet, en théorie, d'évaluer l'impact potentiel d'une action sur l'utilité totale, et ainsi de choisir parmi différentes alternatives l'action la plus à-même d'augmenter la quantité nette de plaisir, chez l'agent mais également pour toutes les autres personnes potentiellement impactées.

Le calcul de félicité est composé de sept critères. Les six premiers présentent les différentes propriétés relatives à la production d'utilité. Ces critères sont (1) *l'intensité*, (2) la *durée*, (3) la *certitude*, (4) la *proximité*, (5) la *fécondité* et (6) la *pureté*. On dira ainsi que la quantité de plaisir que l'on espère tirer d'une activité sportive dépend de son intensité, de sa durée, de la probabilité de sa réalisation, de son éloignement dans le temps, des plaisirs qu'elle induira par la suite et des souffrances qui l'accompagneront. Le septième et dernier critère, celui de *l'étendue*, consiste à répéter le calcul précédent pour toutes les personnes

potentiellement impactées par notre action. Dans le cas de l'activité sportive, on peut raisonnablement penser que l'augmentation de notre espérance de vie induite par une activité sportive régulière, sera appréciée par notre famille et par nos amis, et qu'une plus faible probabilité de développer de lourdes et coûteuses pathologies sera appréciée par tous les contribuables. Dans ce cas-là, l'étendue confirme et accentue la tendance annoncée par les six premiers critères. Il existe de nombreux autres cas pour lesquels la recherche de plaisir immédiat induit par une activité personnelle (par exemple le fait de fumer ou de régulièrement boire de l'alcool), est tempérée par ses potentiels effets indésirables (cancers), voire totalement remise en cause par sa désutilité sociale.

Le calcul de félicité, tel qu'il est présenté ici, présente deux difficultés majeures, [Cléro, 2015 ; Postema, 2006 ; Quinn, 2014]. Ce calcul présuppose, tout d'abord, que toutes les sensations d'un agent puissent être évaluées dans une même métrique. Autrement dit, le calcul de félicité stipule qu'il existe une mesure commune à toutes les sensations quelles que soient leur nature, leur intensité et leurs propriétés. Or, il n'est pas évident qu'une quantité de souffrance donnée puisse être équivalente à une (autre) quantité de plaisir. De même, il n'est pas évident qu'un plaisir (ou une souffrance) physique, puisse être comparé à un plaisir (ou à une souffrance) émotionnelle ou intellectuelle. Il exige, ensuite, que l'agent puisse évaluer, comparer et agréger ce que ressentent les autres personnes. Bentham reconnaît ces difficultés. Il reconnaît également que la variété des personnalités, mais aussi des circonstances dans lesquelles un évènement a lieu, rend quasiment impossible la mesure de son impact sur les agents [Warke, 2000]. Bentham a, également beaucoup varié sur la possibilité de faire de la monnaie l'unité de mesure des plaisirs et des peines [Quinn, 2014]. Cela explique que des commentateurs, tels Mitchell ou Postema, aient préféré voir en Bentham plutôt un classificateur qu'un calculateur, « plus près d'être le Linnaeus que le Newton du monde moral » [Mitchell, 1918, p.182]. Certains, tels Quinn, considèrent que la difficulté pratique soulevée par le calcul de félicité ne remet pas fondamentalement en cause la théorie de Bentham [Quinn, 2014]. D'autres, enfin, adoptent une approche hybride, et voient dans les tentatives de classification de Bentham une autre forme de calcul [Sigot, 2001].

## L'utilitarisme libéral de John Stuart Mill

*John Stuart Mill (1806-1873), figure intellectuelle du XIX<sup>e</sup> siècle*

John Stuart Mill est l'un des intellectuels les plus influents du dix-neuvième siècle. Son œuvre est toujours considérée comme une référence du libéralisme et de l'utilitarisme. Il est également considéré comme l'une des plus importantes figures de l'économie politique classique avec Ricardo et l'un des premiers penseurs à avoir ouvertement défendu le droit de vote des femmes.

Son père James, historien, économiste et philosophe utilitariste, ami et disciple de Bentham, fut son seul pédagogue. Il accorda un grand soin à l'éducation de son fils, méticuleusement planifiée et organisée selon les principes de la philosophie associationniste et dénuée de toute instruction religieuse. En 1826, alors âgé de vingt ans, Mill traverse une crise psychique (on parlerait aujourd'hui de dépression nerveuse) qui l'amène à reconsidérer ses positions philosophiques ainsi que son allégeance intellectuelle à Bentham et à se rapprocher des Romantiques [Capaldi, 2004]. Il découvre en particulier l'œuvre de William Wordsworth qui l'amène à remettre en cause la suprématie de la raison sur l'imagination. Il découvre également Samuel Taylor Coleridge, qu'il considère comme étant le plus grand esprit de son époque (avec Bentham), et auquel il doit certaines de ses positions les plus conservatrices, telles que l'importance des traditions ou du rôle civilisateur de l'élite intellectuelle (*clerisy*).

Le succès du *System of Logic* en 1846, puis celui des *Principles of Political Economy*, deux ans plus tard, fit de John Stuart Mill une figure intellectuelle publique. Il utilise alors sa notoriété pour défendre ses principes auprès du public. En 1849, il publie dans le *Fraser's Magazine* une cinglante réponse à Carlyle, qui avait défendu dans un précédent numéro l'esclavage. Il prend également position en 1861 pour les nordistes abolitionnistes lors de la guerre civile aux Etats-Unis [Capaldi, 2004, p.306]. A l'inverse, il prend la défense de la politique impérialiste de l'Angleterre dans un essai intitulé *A Few Words on Non-Intervention*, également publié dans le *Fraser's Magazine*. Selon Mill, les nations ont – comme les individus – le devoir moral de soutenir les autres dans leur effort de civilisation. C'est dans ce même magazine, enfin, qu'il publie en 1861 – en trois livraisons – son fameux

essai sur l'utilitarisme, *Utilitarianism*. L'ensemble fut ensuite republié sous forme d'ouvrage en 1863.

Refusant de prêter serment à l'Église anglicane, il ne put enseigner dans une université britannique. Il travailla au sein de la Compagnie des Indes Orientales, et se retirant, en 1858, après trente-cinq ans de carrière. Il se réorienta ensuite vers la politique et fut élu député au parlement de 1865 à 1868. Il défendit, en vain, le système de représentation proportionnel proposé par Hare et l'extension du droit de vote aux femmes. Il meurt à Avignon le 7 mai 1873, où il est enterré aux côtés de son épouse, Harriet Taylor (cf encadré 1).

#### Encadré 1: John Stuart Mill et Harriet Taylor

Harriet Taylor (1807-1858) a eu une importante influence intellectuelle sur Mill. Celui-ci la rencontra en 1830. Tous deux fréquentent le cercle de William Johnson Fox, un penseur politique et religieux unitarien proche des philosophes radicaux. Très influencée par la pensée des Lumières, Harriet convainc Mill de rédiger un petit opuscule, *Du génie*, dans lequel ce dernier tente de concilier la raison des Lumières avec l'imagination des romantiques (Capaldi, 2004: 104). Harriet est alors mariée à Henry Taylor. Bien que très épris l'un de l'autre, ils attendront la mort de celui-ci pour se marier en 1851. Celle-ci tomba, cependant, rapidement malade et malgré des séjours prolongés dans le sud de la France, elle mourut en 1858. Mill acheta une maison à Avignon, où Harriet avait été enterrée, et y resta dès lors six mois par an. Il y mourut, et y fut également enterré, en 1873.

Mill fait un portrait dithyrambique de son épouse dans son *Autobiography*, et lui attribue tout le mérite de ses propres écrits. Il est difficile d'établir avec précision le rôle exact que joua Harriet dans la genèse de son œuvre. Elle semble néanmoins avoir joué un rôle déterminant dans au moins trois écrits, les *Principles of Political Economy* (1848), *On Liberty* (1859) et *Subjection of Women* (1869). D'après son autobiographie, c'est Harriet Taylor qui l'a convaincu de la nécessité d'un chapitre sur le futur probable des classes laborieuses dans ses *Principles* (partie IV, chapitre 7). Friedrich Hayek lui reprocha d'ailleurs d'avoir perverti le libéralisme de Mill par ses positions socialistes (Hayek, 1969). Il est difficile d'établir l'influence de Harriet sur *On Liberty*, premier ouvrage à paraître après sa mort. Mill présente l'ouvrage à sa publication comme une production commune, mais peu d'éléments viennent étayer ou infirmer cette thèse. En revanche, l'influence de Harriet Taylor est incontestable dans *Subjection of Women*. Celle-ci avait, en effet, déjà publié un essai sur le vote des femmes ('The Enfranchisement of Women') en 1851 dans la *Westminster Review*, essai qui rejoint dans ses grandes lignes les arguments avancés dans *Subjection of Women*.

### *L'hédonisme qualitatif de Mill*

L'utilitarisme de Mill, comme celui de Bentham, repose sur le rejet de l'intuitionnisme et des autres théories aprioristes de l'éthique, dont la théologie. La position épistémologique de Mill est ancienne, elle a été développée en particulier dans un traité, *System of Logic*, publié en 1843, dans lequel il s'attaquait à la doctrine de William Whewell, qui essayait alors de trouver une voie moyenne entre l'apriorisme kantien et le sensualisme lockien. Selon Whewell, il est possible aux hommes de découvrir, par une méthode appropriée, les 'vérités nécessaires'. Mill, qui voit dans la philosophie intuitionniste de Whewell un bastion du conservatisme moral [Ryan, 1974, p.60], adopte, en revanche, une démarche empiriste et associationniste [Skorupski, 1989]. L'associationnisme est une théorie psychologique, héritée de Locke et de Hume, selon laquelle le processus de pensée se fait par association d'idées. D'un point de vue épistémologique, cela signifie que nos connaissances ne sont ni innées ni intuitives mais qu'elles sont le produit d'une chaîne cognitive qui prend source dans nos sensations. Chez Bentham et chez James Mill, l'associationnisme justifie la passivité des individus, dominés par leurs désirs. John Stuart Mill fait cependant de l'hédonisme, et de l'associationnisme qui le sous-tend, un outil du progrès de la nature humaine.

L'utilitarisme de Mill se distingue, en revanche, de celui de Bentham sur au moins un aspect essentiel : son approche de l'hédonisme. A la différence de Bentham, Mill considère que la mesure des plaisirs ne peut se réduire à sa seule quantité. Il introduit donc une dimension qualitative, en parlant notamment de plaisirs 'nobles' (*higher pleasures*), répondant ainsi à Carlyle, qui avait accusé l'utilitarisme de ne voir dans l'homme qu'un animal en quête de vils plaisirs, et qui l'avait qualifié de 'philosophie de cochons' (*philosophy of swines*). La théorie de Mill, connue aussi sous le nom d'hédonisme qualitatif, manque de clarté et fait l'objet de nombreux débats entre commentateurs.

L'introduction de la qualité, en tant que propriété distincte de la quantité, dans le schéma utilitariste peut être interprétée de trois manières. La première voit dans celle-ci une rupture avec le dogme de l'hédonisme [Riley, 1999, 2003]. Si, la qualité du plaisir induit par certaines activités (par exemple la poésie) est lexicalement supérieure à celui induit par d'autres (par exemple le rap), alors aucune quantité de rap ne permettra jamais d'égaliser le

plaisir procuré par la poésie. Une seconde interprétation considère que si la poésie est (hypothétiquement) toujours préférable au rap, ce n'est pas parce qu'elle suscite des sensations d'un type différent mais que celles-ci sont plus intenses, autrement dit qu'elles procurent du plaisir en plus grande quantité [Crisp 2002, 2006]. Selon cette approche, la dimension qualitative de l'évaluation plaisir est réductible à sa dimension quantitative, et elle ne remet donc pas en cause le principe de l'hédonisme. Il existe, enfin, une troisième façon d'aborder le texte de Mill. Celle-ci consiste à considérer que la qualité et la quantité de plaisir déterminent conjointement l'utilité finale de ladite activité sans qu'il soit possible d'identifier l'importance relative de l'une ou de l'autre [Donner 1998 ; Saunders 2016].

### *L'utilitarisme libéral de Mill*

En dépit de son utilitarisme, Mill est aujourd'hui considéré comme l'une des principales figures du libéralisme moderne. Le principe de non-nuisance (*harm principle*), qu'il développe dans *On Liberty* (1859), affirme que la seule raison légitime pour une société de contraindre un individu est de l'empêcher de nuire aux autres. Toute politique consistant à prévenir ses actions pour d'autres raisons, par exemple parce qu'on les considère contraire à son propre bien, est par conséquent illégitime. Ce principe constitue l'un des piliers de la tradition libérale moderne. Son articulation avec la philosophie utilitariste de Mill pose, cependant, question.

Une première position possible consiste à considérer l'utilitarisme et le libéralisme comme deux positions incompatibles. La maximisation de l'utilité totale exige, en effet, parfois, de sacrifier celle d'individus ou de minorités. Cela va non seulement à l'encontre du principe de non-nuisance mais cela constitue également une violation des droits élémentaires des individus. Le libéralisme contemporain, tel qu'il est représenté par Rawls ou Dworkin, par exemple, est clairement anti-utilitariste. C'est la position défendue par Isaiah Berlin, John Gray et C.L. Ten, notamment.

Selon ces auteurs, le libéralisme de Mill prévaut sur son utilitarisme. Ils rejettent, en particulier, l'idée souvent avancée que le principe de non-nuisance se justifierait par le principe de maximisation de l'utilité (Berlin 1969 ; Ten 1980). Gray récuse également l'idée

selon laquelle Mill défendrait un libéralisme perfectionniste, dans lequel la supériorité de certains plaisirs, ou de certains styles de vie, sur d'autres, ne constituerait pas une contrainte sur la liberté des individus à s'auto-déterminer [Gray, 1991]. Tous considèrent, enfin, l'expression utilitarisme libéral comme trompeuse.

A l'inverse, d'autres commentateurs considèrent que le projet utilitariste est tout à fait compatible, voire indissociable, avec un engagement libéral [Brink 2013, 2018 ; Riley, 1998]. Jonathan Riley affirme ainsi que le principe d'individualité, sur lequel repose sa doctrine de l'utilité, est coextensive de sa théorie de l'utilité [Riley, 1998, p.159]. Plus récemment, David Brink a développé la thèse selon laquelle un utilitarisme indirect (autrement dit un utilitarisme de la règle, cf chapitre 3), suffisait seul à justifier les droits des individus chez Mill. Ce qu'il appelle une 'conception utilitariste-sanction des droits' (*sanction-utilitarian conception of rights*) [Brink, 2013, p.218].

La question de l'articulation entre l'utilitarisme et le libéralisme chez Mill dépend, certes des éléments textuels qui peuvent parfois prêter à confusion, mais également de la signification que l'on donne à ces deux termes. L'utilitarisme, on le verra, n'est pas un système de pensée homogène. Le fait que l'on interprète la pensée de Mill comme appartenant au groupe des théories de l'utilitarisme de l'acte ou de la règle modifie profondément le rapport que sa théorie entretient avec le libéralisme. De la même façon, les débats sur la nature plus ou moins élitiste ou perfectionniste de sa théorie de la valeur influencent énormément l'appréciation de ses arguments sur la liberté ou l'autonomie des individus. Mais il est fort possible, également, que notre traitement de la question soit biaisé par une conception moderne (qu'elle s'inspire de Berlin, de Hayek ou de Rawls) de ce qu'est le libéralisme. Or, le libéralisme ne se réduit ni au pluralisme des valeurs, ni à un ordre spontané ni au contractualisme [Audard, 2011 ; Ryan, 1998]. A l'époque de Mill, les valeurs du libéralisme, celui de Humboldt, de Greene ou de Dewey, étaient celles de l'autonomie et du développement de soi, et non celles de la protection des droits individuels. Cette mise en perspective permet, par ailleurs, de mieux comprendre les positions parfois surprenantes de Mill à l'égard du paternalisme, pourtant a priori rejeté par le principe de non-nuisance [Clayes, 2013], ou encore à l'égard des colonies anglaises [Bell, 2010 ; Schultz et Varouxakis, 2005].

## **Les Méthodes de l'éthique d'Henry Sidgwick**

*Henry Sidgwick (1838-1900), le professeur de Cambridge*

A la différence de Bentham et de Mill, Henri Sidgwick fut, avant tout, un enseignant et un universitaire. Après de brillantes études à Rugby, Sidgwick est admis à Cambridge, au Trinity College, en 1856. Il y passera l'essentiel de sa vie. Il démissionna néanmoins de sa position de *fellow* à Trinity en 1869 pour manifester son opposition à l'obligation faite aux enseignants de se soumettre à la doctrine de l'Eglise anglicane [Schultz, 2004, p.124]. Sidgwick a activement soutenu l'accès des femmes à l'université et il joua avec son épouse, Eleanor Mildred Balfour, un rôle déterminant dans la création de Newnham College, premier *college* réservé aux filles à Cambridge.

Sidgwick connut un succès immédiat avec son premier ouvrage, les *Methods of Ethics* (1874), qui connurent sept éditions successives. Professeur de philosophie morale, il s'intéressa également beaucoup à l'économie et à la politique. Ses *Principles of Political Economy* (1883), aujourd'hui oubliés, eurent une influence importante, particulièrement auprès d'Alfred Marshall, et ce malgré de nombreux points de désaccords [Backhouse, 2006]. Très influencées par sa philosophie utilitariste, ses positions politiques semblent parfois manquer de cohérence, et oscillent entre réformisme libéral et conservatisme impérialiste [Schultz, 2007].

John Rawls a peut-être eu tort de présenter Sidgwick, dans sa *Théorie de la justice*, comme l'un des représentants de l'utilitarisme classique avec Bentham et Edgeworth [Rawls, 1997, p.58-59]. Même s'il s'inscrit dans un cadre de réflexion qui porte fortement l'empreinte des penseurs utilitaristes qui l'ont précédé, Sidgwick se démarque d'eux dans la forme comme dans le fond. Dans la forme, parce qu'à la différence de Bentham (et dans une moindre mesure de Mill), Sidgwick n'a pas la posture d'un chef d'école de pensée. Il n'est ni dogmatique ni réformateur. Il se cantonne dans la position du savant, intéressé par la seule recherche de la 'vérité' morale. Celle-ci pouvant même, comme c'est le cas de la première édition des *Méthodes*, se révéler brutalement déstabilisante (et par conséquent inacceptable) pour l'utilitariste convaincu. Dans le fond, ensuite, parce que Sidgwick ne



partage pas certains des fondamentaux de l'utilitarisme 'classique', tel que l'hédonisme psychologique ou le mépris du sens commun. Schneewind a donc probablement raison de préciser qu'il s'agit d'une erreur de considérer les *Méthodes de l'Éthique* avant tout comme une défense de l'utilitarisme [Schneewind, 1977, p.192].

Contrairement à Mill, Sidgwick est profondément influencé par Kant [De Lazari-Radek et Singer, 2014]. La philosophie morale de Sidgwick est construite sur deux principes 'évidents', immédiatement appréhendables par l'intuition, le principe de prudence – sur lequel repose l'hédonisme égoïste – et la bienveillance rationnelle sur laquelle repose l'utilitarisme. On reconnaît ici l'influence de Kant, auquel Sidgwick rend d'ailleurs hommage [Sidgwick, 1874 réed. 1981, p.386], et de son principe d'universalisation. Appliqué à la recherche du bien, ce principe affirme, en effet, qu'un homme rationnel n'a pas davantage de raison de privilégier le bien-être d'un homme à celui d'un autre, et par conséquent, de privilégier son propre bien-être à celui d'autrui.

Il est également important, pour bien comprendre la démarche de Sidgwick, de garder à l'esprit le contexte intellectuel dans lequel il écrivait [Singer, 1992]. Celui-ci est doublement marqué par le positivisme dominant et le darwinisme naissant des années 1860. Si Sidgwick a relativement ignoré le second mouvement, il s'inscrit en revanche dans une approche positiviste inspirée en partie par Comte, Seeley et Renan [Schultz, 2004, p.81]. Sidgwick est animé par la conviction qu'il est possible d'établir de façon rigoureuse (et donc scientifique) les fondements de la morale sans faire appel à la religion. C'est là la fonction qu'il assigne à l'éthique. Mais qui dit démarche scientifique, dit en général unité de discours ou, du moins, de méthode.

### *Les trois méthodes de l'éthique*

Sidgwick entreprend l'étude de ce qu'il considère être les trois principales méthodes de l'éthique : l'égoïsme, l'intuitionnisme et l'utilitarisme. Il est important de souligner que l'objet des méthodes de l'éthique est de déterminer ce que l'on *doit* finalement faire dans telle ou telle situation, et non simplement de nous donner des raisons d'agir. Or, pour Sidgwick, ce que l'on doit faire signifie : ce que la raison nous impose de faire. Le caractère

prescriptif de la morale est donc, pour Sidgwick comme pour la plupart des utilitaristes, une conséquence du caractère prescriptif de la raison.

Une méthode de l'éthique est une procédure rationnelle grâce à laquelle il est possible de déterminer le devoir moral de l'homme. Et par devoir, il entend ce qui est absolument commandé par la raison (*'categorically prescribed by reason'*). Il existe un certain nombre d'arguments rationnels en faveur de l'égoïsme et l'utilitarisme qui permettent de justifier dans un cas, le souci de soi, et dans l'autre le sacrifice de soi. Dans le cas de l'égoïsme, la préférence de soi sur les autres peut notamment se justifier par la distinction (ou ce que l'on appelle aujourd'hui la séparabilité) des personnes. Le fait, empiriquement vérifiable, qu'il existe une différence entre moi et les autres, de telle sorte qu'il est rationnel pour moi d'agir pour mon propre bien plutôt que pour celui des autres. Dans le cas de l'utilitarisme, il est également possible de démontrer – en partant du principe d'impartialité – que, du point de vue de l'univers, mon propre bonheur ne compte pas davantage que celui des autres, et donc que ce qui est rationnel de faire est de rechercher le maximum de bonheur, quels qu'en soient ses bénéficiaires. La séparabilité et l'impartialité sont considérés comme indémontrables. La finalité de l'action est rationnelle, et a donc une valeur prescriptive, lorsqu'elle est le résultat logique des principes de séparabilité ou d'impartialité.

La troisième méthode de l'éthique est l'intuitionnisme. Influencé par sa relecture de Joseph Butler (1692-1752), Sidgwick est amené à réévaluer ses positions sur l'intuitionnisme qu'il avait dans un premier temps – à l'instar de Bentham et Mill – rejeté comme vague et arbitraire. Il distingue l'intuitionnisme moral (ou sens commun) de l'intuitionnisme philosophique. L'intuitionnisme moral ou le sens commun est la doctrine, attribuée à William Whewell (bien que cette interprétation soit remise en cause [Schneewind 1974]), selon laquelle la norme morale s'imposerait spontanément à l'homme sans que ce dernier ait besoin de faire appel à sa raison. L'intuitionnisme philosophique, en revanche, est une méthode rationnelle d'investigation de l'éthique qui consiste à faire de l'évidence un véritable principe de démonstration.

Selon Sidgwick, l'intuitionnisme comme méthode éthique doit répondre à deux exigences pour être valide : 1) il doit pouvoir énoncer avec précision et clarté des principes premiers qui vont de soi et 2) il doit montrer que ces principes sont bien ceux qui sont communément

utilisés dans la pratique [Sidgwick, 1874, réed. 1981, p.343]. Quatre conditions doivent être réunies pour satisfaire la première exigence. Tout d'abord, la proposition doit être énoncée en termes clairs et précis. C'est ce que Sidgwick appelle le critère cartésien. Deuxièmement, le caractère évident de la proposition doit être établi par un raisonnement rigoureux. Troisièmement, les propositions reconnues comme évidentes doivent être cohérentes entre elles. Enfin, les propositions que je considère comme évidentes doivent également être reconnues comme telles par les autres. Sidgwick n'a pas eu à tester l'aptitude de l'intuitionnisme à répondre à ces quatre conditions: selon lui, l'intuitionnisme ne remplit pas le premier critère d'une véritable méthode éthique, à savoir le critère cartésien [Sidgwick, 1874, réed. 1981, p.342-43]. L'intuitionnisme moral, tel qu'il se présente chez les moralistes tels Whewell et Grote, doit donc être abandonné.

L'intuitionnisme philosophique, en revanche, respecte ces conditions. La rationalité d'une action y est donnée comme évidente en soi. Celle-ci partage également avec l'utilitarisme le bienfondé du principe d'impartialité. Il n'est donc pas étonnant que l'intuitionnisme et l'utilitarisme convergent *in fine* [Sidgwick, 1874, réed. 1981, p.496]. Encore cela demande-t-il à Sidgwick de parfois gommer les fondements égoïstes de certaines intuitions, comme c'est le cas de la prudence notamment. Cette réhabilitation de l'intuitionnisme comme méthode de l'éthique, et la conclusion selon laquelle nos principes intuitifs convergent en pratique avec ceux de l'utilitarisme, constitue l'une des thèses les plus fortes des *Méthodes*.

### *Le dualisme de la raison pratique*

Ayant démontré que sur les trois méthodes de l'éthique, l'une d'entre elles (l'utilitarisme) peut être présentée comme la forme délibérée et scientifique de l'autre (l'intuitionnisme), il ne reste que deux méthodes éthiques, l'utilitarisme et l'égoïsme, à départager. Or il ne peut y avoir deux procédures de décision rationnelles justifiant deux types d'action totalement opposées. C'est, en effet, « un postulat fondamental de l'éthique, que, dans la mesure où deux méthodes sont en conflit, l'une ou l'autre d'entre elles doit être modifiée ou rejetée » [Sidgwick, 1874, traduit par Terestchenko, 2004, p.104]. Sidgwick se trouve donc confronté à ce qu'il appelle lui-même un dualisme de la raison pratique et qu'il considère être le problème fondamental de l'éthique.

Réconcilier la recherche de l'intérêt personnel avec celle de l'intérêt collectif constitue, on le sait, l'enjeu principal de l'utilitarisme, et c'est généralement à la manière dont les philosophes ont répondu à cette question qu'est jugée la valeur de leur pensée. Pour Bentham, c'est aux législateurs de faire en sorte que ces deux objectifs coïncident. Pour Mill, c'est de la sanction intérieure, la conscience morale, que l'on est en droit d'attendre cette réconciliation. Sidgwick aborde la question d'un strict point de vue analytique. Il ne s'intéresse pas aux différentes sanctions auxquelles pourrait être soumise la recherche de l'intérêt personnel. Il recherche une solution théorique au problème. Or, tel qu'il se présente le problème est insoluble, à moins de supposer l'existence d'un 'gouvernement moral de l'univers'.

La conclusion de la première édition des *Méthodes* (1874) est on ne peut plus pessimiste. Sidgwick finit, en effet, par admettre que « le cosmos du devoir est ainsi réellement réduit à un chaos, et l'effort prolongé de l'intellect humain pour former un idéal parfait de la conduite rationnelle est perçu comme ayant été d'avance condamné à un échec inévitable » [Sidgwick, 1874 traduit par Terestchenko, 2004, p.124]. Cette dernière phrase est supprimée des éditions ultérieures, mais rien ne semble démontrer une évolution de sa part [Martinoia, 2011 ; Terestchenko, 2004]. Le dualisme de la raison pratique semble donc être irréductible. Il n'existe pas de procédure rationnelle permettant d'arbitrer entre le devoir que l'on a envers soi et le devoir que l'on a envers les autres.

Le dualisme de la raison pratique, et ses possibles interprétations, occupe une place fondamentale dans les études sidgwickiennes. Si certains auteurs, tels Mackie, admettent l'impossibilité de prouver la supériorité des raisons utilitaristes sur les raisons égoïstes, voire l'impossibilité d'établir une théorie éthique sur des fondements rationnels [Mackie, 1976], d'autres préfèrent proposer des solutions possibles au dualisme qu'ils espèrent être conformes, sinon à la lettre, du moins à l'esprit de l'œuvre du philosophe de Cambridge [Brink, 1988 ; Crisp, 2015 ; Frankena 1974, 1992 ; Orsi, 2008 ; Phillips, 2011, Skorupski, 2001].

## Conclusion

Bentham, Mill et Sidgwick sont reconnus comme les principaux fondateurs de l'utilitarisme 'classique'. Mais, bien que leurs théories s'appuient sur un socle commun, celles-ci diffèrent considérablement. Dans quelle mesure peut-on alors parler d'utilitarisme ? Quel est le socle commun sur lequel s'appuient ces théories ?

Le premier trait caractéristique de l'utilitarisme, tel qu'il apparaît chez ces trois auteurs, est sa posture hédoniste. Le jugement moral des actions est déterminé par la somme des plaisirs et des peines provoqués par ladite action. Mill développe une théorie qualitative de l'hédonisme que certains considèrent, cependant, comme incompatible avec le calcul de félicité benthamien.

Ces théories ont, ensuite, en commun le fait qu'elles traitent tous les individus comme égaux. Chez Bentham, notamment, tous les individus comptent pour un et seulement pour un. Bien que Mill approuve également ce principe, certaines de ses positions politiques, notamment le vote pluriel ou son soutien au colonialisme, ont pu être perçues comme des entorses à celui-ci malgré son combat sincère pour l'égalité des femmes et des hommes. Davantage qu'une position envers l'égalité sociale ou politique entre les individus, Sidgwick défendra un principe d'impartialité, inspiré de Kant, dans lequel tous les individus sont identiques, et donc substituables, du point de vue de l'Univers.

Il n'existe donc pas, à proprement parler de théorie classique de l'utilitarisme, au sens où il existerait un dogme unique, une théorie orthodoxe, à laquelle on puisse se référer. Il existe des théories historiques, qui sont sujettes à des interprétations variées et contradictoires. On peut, cependant, admettre qu'il existe un 'esprit' utilitariste qui unit ces différents auteurs, ainsi que leurs successeurs. Cette commune attitude a été désignée par Thomas Scanlon, l'utilitarisme philosophique (*philosophical utilitarianism*) et elle se réduit à deux principes. Le premier est que le seul fait moralement important concernant les individus est leur bien-être. Le second est que le bien-être d'un individu équivaut à celui de n'importe quel autre (principe d'impartialité). Les fondements et implications de ce principe seront traités dans le chapitre 4. Le prochain chapitre présente les différentes approches du bien-être qui sont actuellement utilisées par les philosophes utilitaristes.

## II/ MESURER, AGREGER ET MAXIMISER L'UTILITE

L'utilitarisme classique est une philosophie qui permet de juger de la légitimité morale d'une action au regard de ses conséquences sur la somme nette totale nette des plaisirs. Il s'agit, ainsi que nous l'avons vu dans le premier chapitre, d'une théorie hédoniste. L'utilitarisme contemporain se distingue cependant de l'utilitarisme classique sur (au moins) deux aspects : l'abandon de l'hédonisme et l'adoption d'une approche indirecte du conséquentialisme. Ce chapitre est consacré à l'abandon de l'hédonisme et à ses implications théoriques. Il présente les différentes théories de la valeur associées à l'utilitarisme ainsi que les différents débats, parfois techniques, soulevés par la mesure, l'agrégation et la maximisation de l'utilité.

### **Les trois conceptions du bien-être**

On doit à Derek Parfit la classification des théories du bien-être en trois grandes approches : l'hédonisme, la satisfaction des désirs (aussi nommée théorie du succès ou des états du monde) et les théories de la liste objective [Parfit, 1984, p.4]. La théorie hédoniste désigne l'approche classique, celle notamment développée par Bentham, mais également un certain nombre de variantes plus contemporaines. Elle est également connue aujourd'hui sous le nom de théorie des états mentaux. Adopter une approche plutôt qu'une autre a des implications majeures sur le calcul d'utilité et peut radicalement changer les conclusions d'un même raisonnement utilitariste. Historiquement, les théories utilitaristes ont beaucoup évolué, abandonnant progressivement la notion d'état mental au profit de celle d'état du monde, mais on observe depuis un certain nombre d'années déjà un retournement de tendance.

### *La théorie des états mentaux*

L'approche des états mentaux regroupe un certain nombre de théories qui ont toutes en commun le fait de considérer le bien-être individuel du point de vue de l'individu. Ces théories se donnent généralement pour but d'interroger le ressenti ou l'attitude de l'individu pour évaluer l'utilité d'un bien ou d'un évènement. Cela signifie, notamment, que contrairement à ce qu'affirme G.E. Moore, un évènement ne peut être qualifié de plaisant ou de désirable que s'il existe un individu pour l'apprécier en tant que tel. Cela ne signifie pas, néanmoins, que la théorie des états mentaux soit purement subjective ou qu'elle ne soit appréhendable que par introspection. L'hédonisme benthamien en est un bon exemple. Chez Bentham, bien que l'utilité exprime un état ou un ressenti de l'individu, celle-ci dépend néanmoins des propriétés réelles du bien ou de l'évènement en cause. Le plaisir, ou la souffrance, ont une réalité objective, et celle-ci peut être isolée. On a ainsi vu dans le précédent chapitre qu'il était théoriquement possible pour Bentham d'en déterminer la pureté, l'intensité ou la fécondité. D'autres auteurs préfèrent voir dans le bonheur des individus un état psychologique ou émotionnel qui leur est strictement personnel et subjectif.

On considère généralement qu'il existe deux approches possibles de la théorie des états mentaux, l'approche *internaliste*, selon laquelle le caractère plaisant ou déplaisant d'une expérience relève d'une qualité qui lui est propre (la qualité d'être 'plaisant'), et l'approche *externaliste*, selon laquelle une expérience peut être considérée plaisante dès lors qu'elle est identifiée en tant que telle par les agents [Sumner, 1996; Haybron 2016]. On distingue parfois un troisième type d'approche, dite d'attitude ('attitudinale'), lorsque celle-ci n'est pas associée à une variante de l'externalisme, et selon laquelle le plaisir désigne une attitude, un état de conscience, vis-à-vis d'une situation que l'on *juge* plaisante [Feldman 2004].

L'hédonisme de Bentham et de Mill appartient à la première catégorie. Mill et Bentham sont partisans de l'hédonisme psychologique, c'est-à-dire qu'ils posent comme un fait empirique le fait que les êtres humains sont mus par la recherche de leur plaisir. Selon Bentham, la sensation de plaisir est causée par une qualité intrinsèque du bien, qu'il est

incapable d'identifier, mais qu'il suppose être commune à tous les objets ou événements plaisants.

La thèse internaliste, selon laquelle l'ensemble des expériences plaisantes auraient en commun une propriété, qui leur serait propre, et qui serait la cause de la sensation de plaisir chez l'homme a fait l'objet de sévères critiques [Griffin, 1986]. La grande diversité des expériences plaisantes (pensez par exemple à la différence qui existe entre le fait de manger un gâteau au chocolat, de lire un roman, de vivre une histoire d'amour ou d'admirer un coucher de soleil), ainsi que l'extraordinaire disparité qui existe entre les goûts individuels semble, en effet, discréditer cette hypothèse.

Sidgwick est, quant à lui, un représentant de l'approche externaliste. Dans la mesure où ce ne sont pas les expériences en elles-mêmes qui sont plaisantes mais le fait qu'on les trouve désirables, il est relativement aisé de justifier – sans contrevenir au principe de l'hédonisme quantitatif – que certaines expériences, celles qui impliquent des activités intellectuelles ou artistiques par exemple, sont considérées comme plus désirables que certaines autres. La détermination de l'utilité d'une expérience, ou de l'intensité de son caractère plaisant, est en effet le produit d'une double opération cognitive, celle qui consiste à se représenter la désirabilité de ladite expérience – notamment par comparaison avec d'autres d'expériences – et celle qui consiste à apprécier la désirabilité de l'état mental dans lequel cette expérience me place. Or les deux représentations ne sont pas équivalentes. Cela explique, notamment, que, pour Sidgwick, l'hédonisme ne soit pas incompatible avec la recherche de la vertu. C'est parce que, dit-il, nous considérons comme désirable en soi l'état de conscience qui est le nôtre lorsque nous nous engageons dans des activités vertueuses que nous jugeons celles-ci comme faisant également partie du Bien Ultime [Sidgwick 1874, réédition 1981, p.397].

### *La théorie du succès et l'utilitarisme des préférences*

Une seconde façon de concevoir le bien-être consiste à prendre en compte la satisfaction des désirs ou des préférences des agents. C'est l'approche dite des désirs, ou encore la théorie du succès ou des états du monde. Elle se distingue notamment de l'hédonisme en



ce qu'elle considère l'utilité comme le produit ou l'expression d'un choix plutôt que sa cause ou sa motivation. Contrairement à l'hédonisme, et plus généralement à la théorie des états mentaux, l'utilité d'un bien, d'une action ou d'un événement est directement observable à partir des actions des individus. Très appréciée des économistes, cette théorie de la valeur est à l'origine de l'utilitarisme des préférences, dominant depuis la seconde moitié du XX<sup>e</sup> siècle. Parmi ses plus importants représentants, on trouve notamment John Harsanyi, Richard Hare, Alfred Ayer, Richard Brandt (avant qu'il ne fasse marche arrière) ou, jusqu'à encore très récemment, Peter Singer.

La théorie des désirs a été très en vogue tout au long du XX<sup>e</sup> siècle. Son succès peut être attribué à deux raisons principales. La première est le développement de l'économie du bien-être et le besoin des économistes d'observer et de comparer les utilités des individus. Une seconde raison tient, semble-t-il, à sa bonne adéquation avec l'esprit des sociétés modernes, qui ainsi que le souligne Sumner, aime à voir dans les agents des êtres autonomes, capables de poursuivre des fins qu'ils se sont eux-mêmes données [Sumner, 1996, p.123]. Elle a, cependant, subi un repli ces dernières années, en partie suite aux revers de la nouvelle économie du bien-être [Baujard, 2017] et en partie à cause de la défection de certains de ses importants représentants, Richard Brandt et Peter Singer en particulier.

La théorie du succès ou des états du monde naît de l'insatisfaction profonde suscitée par la dépendance de l'hédonisme vis-à-vis de la psychologie des individus. Or celle-ci, en plus d'être inaccessible à l'observateur extérieur, et parfois à l'agent lui-même, se caractérise par son instabilité et, parfois, son incohérence. A l'inverse, les préférences des individus sont supposées stables et transitives. Dans la mesure où elles satisfont certains axiomes de rationalité (transitivité, réflexivité et complétude), toutes les préférences individuelles sont considérées comme acceptables.

Cette extrême latitude à l'égard des goûts et des désirs individuels pose néanmoins un certain nombre de questions. Toutes les préférences doivent-elles être également satisfaites ? Comment appréhender les désirs contraires à l'intérêt de l'individu qui les exprime ? Quel statut donner aux préférences externes des agents, autrement dit aux préférences qu'ils forment non pas sur leur propre bien-être mais sur celui des autres ?

[Rüegger, 2011] Comment traiter, plus généralement, les préférences (im)morales et/ou (a)sociales des individus, telle l'envie ou le sadisme ?

Face à ces questionnements, un certain nombre d'auteurs ont limité le domaine d'acceptabilité des préférences individuelles. Harsanyi, par exemple, distingue les préférences 'manifestes' des préférences 'véritables' des individus [Harsanyi, 1977]. Les préférences manifestes (*manifest preferences*) d'un agent sont celles qu'il exprime dans la vie courante, c'est-à-dire celles qui peuvent être observées, et qui peuvent par conséquent être fondées sur des croyances ou des faits erronés ou formées sous le coup de l'émotion. Les préférences véritables (*true preferences*) sont définies, au contraire, comme étant celles qu'il aurait eues s'il avait été correctement informé et si celles-ci avaient été formées dans un contexte propice au raisonnement. De la même façon, les désirs informés (*informed desires*) de Griffin [Griffin, 1986], ou les désirs rationnels (*rational desires*) de Brandt [Brandt, 1979], sont les désirs que nous exprimons après réflexion et accès à l'information nécessaire. Les préférences véritables ou les désirs informés ne se prononcent pas sur la rationalité substantielle des préférences. Dans la mesure où les agents sont conscients des conséquences de leurs actions et que leurs désirs reflètent leur personnalité, toutes les préférences individuelles sont légitimes. Contrairement à Harsanyi, Griffin se refuse à restreindre davantage le domaine des préférences possibles, notamment en écartant les préférences immorales.

### *La théorie des listes objectives*

La troisième et dernière grande approche des théories de la valeur est la théorie dite des listes objectives. Comme son nom l'indique, cette théorie considère que l'utilité d'un bien, d'une action ou d'un événement ne dépend pas de l'évaluation qu'en font les agents. L'expression de 'listes objectives' utilisée par Parfit souligne que ces théories se résument bien souvent à une simple liste, plus ou moins exhaustive, de ce que les auteurs considèrent comme objectivement bon. Des trois approches présentées, la théorie des listes objectives est la moins utilisée.

George Edward Moore (1873-1958) est le premier philosophe utilitariste à avoir explicitement rejeter l'hédonisme [Bonicco, 2011]. Il est, avec Hastings Rashdall [Rashdall, 1907; Skelton 2011], le principal représentant de l'*utilitarisme idéal*. Selon cette théorie, une action est bonne (*right*) dans la mesure où elle produit le plus de bien (*good*). Cette proposition est vraie, pour Moore, même si le Bien n'est pas en mesure d'être apprécié. C'est le cas, par exemple, de ce qui est beau. Une action qui produit du beau est bonne (*right*), même si personne n'est en mesure de l'apprécier. C'est du moins la position qu'il adopte en 1903, mais sur laquelle il reviendra ultérieurement [Moore, 1912]. Dans *Ethics*, Moore fait, en effet, de l'état de conscience un élément indispensable de l'évaluation de ce qui est intrinsèquement bon.

La question de la nature intrinsèque du Bien chez Moore est assez complexe et a suscité beaucoup de discussions [Regan, 2003; Bradley, 2006; Hutchinson, 2001]. Retenons simplement que Moore considère l'amour et la beauté comme les deux principales manifestations du Bien, ce qui lui valut un certain succès auprès des membres du groupe dit de Bloomsbury, et notamment de Keynes, de Virginia Woolf et de Lytton Strachey [Levy, 1979; Regan, 2012]. La philosophie non-naturaliste de Moore eut également un impact important dans la première moitié du XX<sup>e</sup> siècle, et notamment auprès des philosophes intuitionnistes H. A. Prichard, C. D. Broad, et W. D. Ross [Hurka, 2003].

La théorie des *capabilities* d'Amartya Sen constitue un autre exemple de conception objective du bien-être, même si elle n'est pas, à strictement parler, utilitariste [Sen, 1979]. Elle s'inspire de la théorie des biens premiers de John Rawls. Rawls définit les biens premiers comme « ce que l'on suppose qu'un être rationnel désirera quels que soient ses autres désirs » [Rawls, 1971, réédition 1987, p.122], autrement dit, les biens dont il préférera toujours avoir plus que moins quels que soient ses projets. Il s'agit généralement de ceux dont la possession assure aux hommes de plus grandes chances de réaliser leurs désirs, progresser dans leurs objectifs et réussir leurs projets. Parmi ceux-ci, Rawls mentionne notamment les droits, les libertés et les possibilités offertes, ainsi que le revenu et la richesse [Rawls, 1971, réédition 1987, p.123]. Dans la mesure où ils sont liés à la structure de base, il s'agit généralement de biens sociaux. Sen considère, cependant, que cette théorie fait peu de cas de la diversité des êtres humains. Les hommes ont, en effet, des besoins très différents selon leur état de santé, leur longévité, les conditions

climatiques dans lesquels ils vivent, leurs conditions de travail, leur tempérament etc. [Sen, 1980, p.215-16].

Sen propose alors une théorie alternative fondée sur les notions de *functionings* et de *capabilities*. La notion de *functioning*, que l'on peut traduire par mode de fonctionnement, désigne ce que l'individu peut faire ou être. Parmi la liste de réalisations possibles, on trouve le fait de vivre longtemps et en bonne santé, le fait de savoir lire et écrire, de participer à la vie politique de son pays etc. Les modes de fonctionnement dépendent des ressources dont l'individu dispose mais elles ne sont pas réductibles à celles-ci. Sen reproche à Rawls de faire de la possession de biens ou de ressources un objet en soi et de tomber ainsi dans une sorte de fétichisme. Les modes de fonctionnement ne donnent de la valeur aux biens que dans la mesure où ils peuvent effectivement être utilisés pour réaliser un objectif. Les *capabilities*, que l'on traduit généralement par capacités ou libertés substantielles, désignent l'opportunité qu'ont les agents de choisir la vie qu'ils souhaitent mener. Les capacités d'une personne dépendent de l'ensemble des modes de fonctionnement auxquels elle peut avoir recours. Le bien-être individuel s'apprécie en fonction de la capacité des individus à mener une vie qu'ils considèrent bonne. Dans la mesure où les modes de fonctionnement paraissent objectivement fondés, on pourrait être conduit à penser que la théorie des capacités appartient à la théorie de la liste objective. Ce n'est pas véritablement le cas puisque la valeur de ces modes de fonctionnement est déterminée à l'aune d'un choix de vie, qui reste chez Sen entièrement subjectif.

### **L'utilité comme mesure du bien**

Le choix d'une théorie de la valeur (états mentaux, préférences, liste objective) a des implications importantes sur la procédure d'évaluation de l'utilité des actions, leur comparaison, leur agrégation et finalement leur maximisation.

Le premier enjeu lié à la maximisation ou, selon les cas, à l'optimisation des utilités est naturellement celui de la mesure. Il existe deux approches très différentes, celle de la théorie cardinale de l'utilité (cardinalisme) et celle de la théorie ordinale de l'utilité (ordinalisme). Le deuxième enjeu est celui du critère de maximisation, qui peut soit être

celui de l'utilité totale, soit être celui de l'utilité moyenne. Enfin se pose la question du principe de maximisation lui-même, qui a été également remis en cause.

### *Théories cardinales et ordinales de l'utilité*

Il existe traditionnellement deux approches pour mesurer l'utilité. La première consiste à mesurer l'intensité de l'utilité. On parle alors d'utilité ou de fonction d'utilité cardinale. La seconde consiste à mesurer son importance relative.

La théorie cardinale de l'utilité, aussi parfois appelée cardinalisme, a pour fonction de quantifier l'intensité de plaisir (ou de déplaisir) éprouvé par l'agent suite à l'action. L'anticipation de cet état de (dé)plaisir, dont l'exactitude dépend des expériences passées de l'individu, constitue la cause de l'action. Cela signifie, par exemple, que si l'utilité pour un individu  $\lambda$  de manger une pomme est de 2 et que l'utilité de manger une barre de chocolat est de 4, alors on peut dire que, pour cet individu, l'intensité de l'utilité du chocolat est le double de celle de la pomme. C'est l'approche retenue par Bentham, et notamment par les premiers économistes marginalistes, Léon Walras, Stanley Jevons ou Carl Menger.

Une seconde approche consiste, au contraire, à ne retenir que le classement des utilités, autrement dit à ne retenir que leur importance relative. On parle alors d'utilité ou de fonction d'utilité ordinale. Ainsi, pour reprendre l'exemple précédent, les théories ordinalistes considéreront que si le chocolat a une utilité plus forte que celle de la pomme, il est néanmoins impossible de chiffrer précisément le ratio qui existe entre elles. On trouve cette conception chez le sociologue et économiste Vilfredo Pareto. Dans le cas d'une théorie ordinale d'utilité, il est par définition impossible de déterminer un maximum d'utilité, mais il est néanmoins possible d'identifier un ou plusieurs optima. L'optimum dit de Pareto désigne les situations dans lesquelles il devient impossible d'améliorer la situation d'au moins un agent sans détériorer celle d'au moins un autre

Les approches ordinales de l'utilité permettent au mieux de déterminer des optima au sens de Pareto, mais elles ne fournissent aucun critère permettant d'arbitrer entre ces optima. Cette position est assumée par les théoriciens de la nouvelle économie du bien-être, qui

refusent d'adopter le point de vue normatif de l'utilitarisme, mais elle explique probablement son déclin aujourd'hui [Baujard, 2017].

## *Encadré 2. Mesure de l'utilité et théories du bien-être*

*Bien qu'elles ne soient pas nécessairement d'essence utilitariste, les théories économiques dites du bien-être se sont construites autour de ces débats sur la mesure de l'utilité. A la 'vieille' économie du bien-être d'Alfred Marshall et d'Arthur Pigou, qui reposait sur la comparaison interpersonnelle des utilités, a succédé, à partir des années 30, la 'nouvelle' économie du bien-être portée par Lionel Robbins, John Hicks et Paul Samuelson. En abandonnant les comparaisons interpersonnelles d'utilité, la nouvelle économie du bien-être réduit considérablement la portée normative de ses conclusions. Au mieux peut-elle espérer déterminer un ensemble de situations paréto-optimales. John Harsanyi a été l'un des rares à résister à ce mouvement et à développer un modèle de bien-être social avec une fonction d'utilité cardinale [Harsanyi 1953, 1955]. Mais le recours à une mesure ordinale de l'utilité permet-il seulement de déterminer un état optimal au sens de Pareto ? Le théorème d'impossibilité d'Arrow démontre le contraire [Arrow, 1951].*

*Dans Social Choice and Individual Values, dont la première édition est publiée en 1951, Kenneth Arrow affirme qu'il est impossible de déterminer une fonction d'utilité collective à partir des préférences ordinales des individus, tout au moins en satisfaisant les quatre conditions suivantes : 1° la condition d'universalité, qui exclut toute restriction sur les préférences individuelles, 2° la condition d'unanimité (ou de Pareto) qui énonce que si tous les individus préfèrent x à y alors la fonction d'utilité collective devra également préférer x à y, 3° la condition d'indépendance des options non pertinentes, qui établit que l'ordre de préférence des options ne peut être affecté par le classement des alternatives non pertinentes et enfin 4° la condition de non-dictature, qui interdit toute imposition d'un classement des préférences individuelles. Face à ce constat, les économistes du choix social ont été amenés à reconsidérer les comparaisons interpersonnelles d'utilité. Ce mouvement, initié par Amartya Sen à la fin des années 70 et bientôt suivi par d'autres économistes tels Hammond, d'Aspremont, Gevers ou encore Blackorby, n'est cependant pas nécessairement synonyme d'un retour au cardinalisme des fondateurs de l'économie du bien-être (et de l'utilitarisme).*

*Dans son ouvrage de 1970, Collective Choice and Social Welfare, Sen réaffecte la fonction normative à la théorie du choix social perdue avec Robbins. Attaché à l'idée de comparaisons interpersonnelles nécessaires au respect des différents principes de justice, il reste néanmoins critique de la théorie d'Harsanyi, dont il remet notamment en cause la portée éthique. Il propose, tout d'abord, de parer à l'arbitraire de la mesure cardinale d'utilité d'Harsanyi en combinant plusieurs mesures cardinales et en ne considérant que les parties communes de leurs classements [Gilardone, 2007, p.224]. Sa réflexion le mènera ensuite à s'interroger sur les fondements de l'échec de l'approche d'Arrow et à identifier sa principale faiblesse dans la pauvreté informationnelle véhiculée par le simple classement des utilités.*

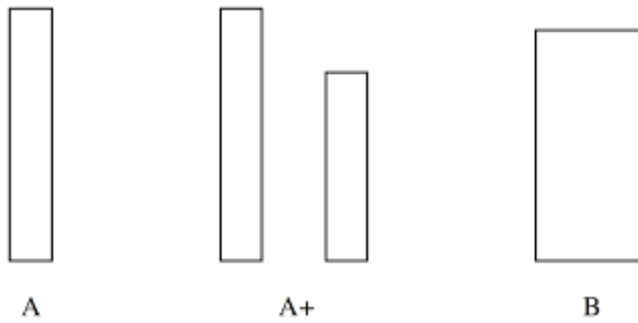
### *Utilité totale ou utilité moyenne ?*

La question de l'importance de la distribution des utilités se pose immédiatement lorsqu'on s'interroge sur l'objet de la maximisation. Sidgwick fut probablement le premier à en voir les implications théoriques [Sidgwick 1874, réédition 1981, p.415-16]. Dans l'utilitarisme classique la variable retenue est celle de l'utilité totale. Mais la maximisation de l'utilité totale pose un certain nombre de problèmes de justice.

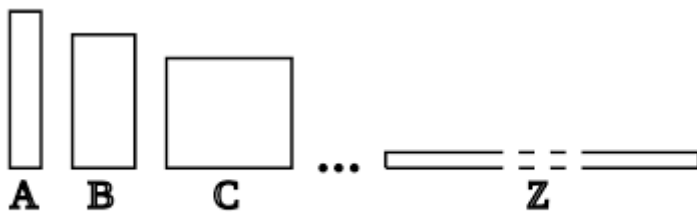
Dans la quatrième partie de *Reasons and Persons*, le philosophe anglais Derek Parfit (1942-2017) dénonce notamment l'absurdité de la maximisation totale de l'utilité qui conduit à préférer des sociétés infiniment peuplées mais dans lesquelles les individus ont une utilité à vivre à peine positive aux sociétés moins peuplées mais composées de gens heureux. C'est l'argument dit de la Conclusion Répugnante [Arrhenius, Ryberg et Tännsjö, 2017].

La conclusion répugnante (*The Repugnant Conclusion*) est un exemple fictif qui a donné naissance à une vaste littérature sur ce qu'on appelle désormais l'éthique de la population [Zuber, 2019]. Elle peut se formuler ainsi: "Pour toute population d'au moins dix milliards d'habitants, ayant tous une très haute qualité de vie, il est possible d'imaginer une population beaucoup plus nombreuse, dont l'existence – toutes choses égales par ailleurs – serait meilleure même si elle n'était composée que d'individus dont la vie vaut tout juste la peine d'être vécue. » [Parfit, 1984, p.388].

Considérons deux populations, A et B. L'abscisse désigne le nombre d'habitants et l'ordonnée le niveau de vie des habitants. Les habitants de A sont moins nombreux que ceux de B mais plus heureux. Ajoutons maintenant à A une seconde population A+ composée d'individus moins heureux. La population totale est alors de B. Son utilité totale (surface) est supérieure à celle de A. Techniquement parlant, B est donc préférable à A.



Son raisonnement est alors le suivant : supposons que l'on continue à ajouter de larges populations à faibles utilités moyennes à des populations plus restreintes mais à forte utilité moyenne. Que se passera-t-il ? L'utilité moyenne de la population ainsi recomposée s'amointrira au fur et à mesure des ajouts mais son utilité totale, elle, augmentera du fait de la taille croissante des nouveaux ajouts. On obtiendra *in fine* une immense population à l'utilité totale maximale mais à la qualité de vie presque nulle. Dans le diagramme ci-dessous, il s'agit de la population Z.



Si l'on suit le raisonnement utilitariste, on devrait conclure que Z est préférable à A. Autrement dit qu'une immense population composée d'individus dont la vie vaut à peine d'être vécue est préférable à une population plus petite composée d'individus qui bénéficient d'une plus grande qualité de vie. Cette conclusion est, pour Parfit, inacceptable. L'objectif de Parfit n'est pas, néanmoins, de proposer un modèle fondé sur l'utilité moyenne qu'il rejette également mais plutôt de conduire à une réflexion sur les problèmes logiques soulevés par la prise en compte le bien-être des générations futures [Parfit, 1986 ; Salvat, 2019a].

Il existe des modèles utilitaristes fondés sur la maximisation de l'utilité moyenne. Le plus connu est certainement celui de l'économiste John Harsanyi (1920-2000). Le modèle de maximisation de l'utilité moyenne d'Harsanyi [1955, 1982] prend appui sur la théorie de l'utilité espérée de Von Neumann et Morgenstern (1944) [Risse, 2002]. Il construit un modèle dans lequel il est possible d'identifier une fonction d'utilité sociale (à maximiser) à



partir des fonctions d'utilités individuelles  $U_i$ . Ces fonctions d'utilité sont cardinales mais également subjectives, ce qui compromet leur agrégation. Harsanyi fait alors appel à un observateur impartial qui préfigure le voile d'ignorance de Rawls. Invité à se placer du point de vue de cet observateur impartial, chaque individu est appelé à ordonner les différents états sociaux possibles. Ignorant tout de ses préférences personnelles, ignorant même jusqu'à son identité, chaque agent est amené à se placer alternativement du point de vue de tous les autres. Il sait, en revanche, que chacun a une probabilité égale d'être l'individu  $i$  dans un état social  $x$ . Harsanyi montre alors que la même fonction d'utilité sociale sera partagée par tous les individus et que celle-ci pourra être définie comme la moyenne arithmétique des utilités subjectives de l'ensemble des membres de la population, c'est à dire  $W(x) = 1/n \sum_i U_i(x)$ . Le principal avantage de l'utilité moyenne est donc de permettre d'intégrer dans le processus d'évaluation un principe de redistribution et de répondre ainsi à l'objection du monstre d'utilité de Nozick (cf Chapitre 4). La maximisation de l'utilité totale peut, dans une société égalitariste, impliquer la maximisation des utilités moyennes. Mais, lorsque ce n'est pas le cas, la question de la priorité de l'une ou de l'autre se pose.

Le modèle d'Harsanyi n'est pas sans rappeler le modèle de Rawls. Bien que proposant une théorie alternative de l'utilitarisme, Rawls considère que seul le point de vue de l'utilité moyenne est conciliable avec la théorie du contrat, qui a ses faveurs. Rawls conteste néanmoins l'hypothèse d'équiprobabilité d'Harsanyi ainsi que la conception minimaliste des personnes qu'impliquerait un raisonnement utilitariste sous voile d'ignorance [Rawls, 1971 trad. 1987]. Le modèle d'Harsanyi est également critiqué pour ne pas être parvenu à démontrer la nature éthique de son raisonnement [Sen, 1977; Roemer, 2008; Weymark, 1991]. Selon ces auteurs, si Harsanyi est bien parvenu à montrer qu'il était possible de sommer les utilités individuelles, cela ne prouve pas qu'il est moralement souhaitable de le faire ; par conséquent, le théorème dit de l'observateur impartial ne peut être présenté comme un argument en faveur de l'utilitarisme. Au-delà des aspects parfois très techniques des arguments présentés, ce débat illustre bien la fracture existant entre l'utilitarisme et le welfarisme, ou entre ce que Sen appelle le conséquentialisme de l'acte (*Act Consequentialism*), qui est une théorie morale, et l'utilitarisme des résultats (*Outcome Utilitarianism*), qui n'est qu'une méthode d'évaluation des états sociaux, elle-même le produit du welfarisme et du classement par sommation (*sum-ranking*) [Sen, 1979].

## Questions relatives à la maximisation d'utilité

La doctrine utilitariste classique, on l'a vu, implique la maximisation de l'utilité ou, dans le cas de fonctions ordinales d'utilité, la recherche d'un état optimal au sens de Pareto. Mais cet impératif de maximisation pose un certain nombre de questions sur la période de référence, sur l'objet de la maximisation (ou minimisation) ou encore sur la surpondération de certaines catégories d'utilités.

### *Sur quel horizon temporel maximiser l'utilité ?*

La première de ces questions concerne la période de référence de la maximisation. Sur quel horizon temporel doit-on maximiser ? La maximisation de l'utilité n'est *a priori* pas limitée dans le temps. Au niveau individuel, on maximise donc sur la vie entière de la personne. Au niveau collectif, sur les vies entières de tous les individus, y compris, des individus encore à naître. Il semble, en effet, difficile de justifier d'un point de vue éthique la maximisation du bonheur des vivants lorsque celle-ci se fait au détriment des générations futures. L'horizon temporel de maximisation de l'utilité est donc, par défaut, celui de l'espèce humaine. Cela soulève, néanmoins, un certain nombre de problèmes, dont celui de l'égalité de traitement de toutes les périodes concernées, autrement dit du principe de neutralité par rapport au temps.

Ce principe a été formulé pour la première fois par Sidgwick [Sidgwick 1874, réédition 1981, p.381] et il est – au moins implicitement – adopté par la plupart des grands théoriciens utilitaristes. Ce principe affirme que le fait qu'un évènement ait lieu plus ou moins loin dans le temps ne constitue pas en soi une raison de modifier notre jugement sur son utilité. La position temporelle de l'évènement ne constitue pas une variable pertinente dans l'appréciation de son utilité. Notons, car c'est un point qui est généralement négligé, que le raisonnement de Sidgwick, qui est souvent sorti de son contexte, est valable lorsqu'on adopte une approche de type hédonisme sophistiqué, comme c'est son cas, ou dans le cas de la théorie des préférences, mais qu'il pose de réelles difficultés dans le cas d'une

approche type états mentaux. Pour David Brink, il se justifie néanmoins dans la mesure où il autorise les compensations inter-temporelles d'utilité [Brink, 2011].

Le principe de neutralité par rapport au temps pose deux difficultés au regard de la maximisation d'utilité. Le premier a trait à la maximisation d'utilité des individus au cours de leur vie et l'autre celui de son articulation avec la maximisation d'utilité entre les générations. Michael Slote soutient que ce principe de neutralité n'est pas compatible avec la réalité de notre cycle de vie [Slote, 1983]. Les hommes et les femmes n'attachent pas la même importance aux événements (qu'ils soient plaisants ou non) qui surviennent dans leur enfance, dans leur vie adulte ou pendant leurs vieux jours. Dans la formulation que lui donne Brink, plus adaptée aux approches des états mentaux, le principe de neutralité par rapport au temps n'est pas incompatible avec une répartition différenciée des utilités au cours du temps. D'autres auteurs opposent à ce principe de neutralité le fait que, dans la réalité, nombreux sont ceux à juger de leur succès à l'aune de la réalisation de leurs projets, du sens qu'ils ont donné à leur vie ou tout simplement de la manière dont celle-ci se termine.

Notons, enfin, la thèse de la non-séparabilité des périodes de la vie d'un individu de John Broome, qui remet *de facto* en cause le principe de neutralité par rapport au temps [Broome, 2006]. Appliqué au niveau intergénérationnel, le principe de neutralité par rapport au temps signifie que l'on doit accorder la même importance au bien-être des individus à naître, y compris dans un lointain futur, qu'à celui des générations présentes. Dans un contexte de raréfaction des ressources naturelles et de réchauffement climatique, cela implique logiquement une réduction du bien-être des générations présentes au profit de celui des générations futures. Selon l'horizon temporel que l'on adopte, la procédure de maximisation de l'utilité peut donc produire des résultats très dissemblables. Notez que le même principe de neutralité par rapport au temps fait appel à des justifications différentes selon qu'il s'applique à un individu ou à une génération d'individus. Dans le cas individuel, les pertes d'utilité dans le présent ou le futur proche sont surcompensées par des gains d'utilité plus importants dans le futur lointain. La neutralité par rapport au temps permet à l'individu de maximiser son utilité au cours de sa vie. Il se justifie donc rationnellement. Dans le cas des générations, les pertes d'utilité subies par les générations présentes ne seront jamais compensées puisqu'elles bénéficieront à d'autres individus dans un futur

lointain. Pour les individus vivants, ce principe ne peut se justifier que d'un point de vue moral.

### *L'utilitarisme négatif*

Le troisième point a trait à l'asymétrie de traitement des souffrances et des plaisirs dans leur comptabilisation. Ce point a été soulevé par Karl Popper en 1952. Selon l'auteur de la *Société ouverte et ses ennemis*, nous sommes – ou nous devrions être – plus sensibles aux souffrances des hommes qu'à leur plaisir et, quitte à accepter un raisonnement conséquentialiste, mieux vaut chercher à minimiser les souffrances qu'à maximiser les plaisirs. Son arrière-pensée est naturellement politique : l'utilitarisme classique justifierait, selon lui, l'établissement d'un despote bienveillant ayant un pouvoir absolu sur la vie privée et intime des individus [Popper, 1952, réédition 2003, p.256]. Ninian Smart (à ne pas confondre avec J.J.C Smart) lui répond que les implications de l'utilitarisme négatif, ainsi qu'il est dénommé, n'en sont pas moins totalitaires [Smart, 1958]. Il lui oppose notamment l'argument selon lequel l'utilitarisme négatif justifierait qu'un dictateur détruise le monde pour mettre fin aux souffrances humaines. Au-delà de ces problématiques politiques, l'utilitarisme négatif pose de réelles questions sur la justification donnée (ou dans ce cas non donnée) au traitement symétrique des souffrances et des plaisirs dans le calcul d'utilité. Existe-t-il, par exemple, un niveau d'utilité symétriquement équivalent à la souffrance provoquée par la torture, par la mort d'un être aimé ou par l'agonie ? Scanlon, par exemple, ne le croit pas. A supposer, cependant, que ce soit le cas, et qu'on trouve une personne jurant que le bonheur qu'elle éprouve à conduire une voiture de luxe est l'exact symétrique du malheur éprouvé par une mère qui perd son enfant, doit-on moralement leur donner le même poids ? Doivent-ils compter de façon égale dans notre tentative de maximisation de l'utilité totale ?

### *Doit-on absolument maximiser l'utilité ?*

Une action doit-elle nécessairement viser le maximum de bien pour être considérée comme morale ? Au-delà des difficultés techniques que soulèvent la maximisation, l'idée

même de juger une action par son importance plutôt que par sa nature pose question. Une action peut être morale sans qu'elle ne soit *la plus* morale possible au même titre qu'une action peut être rationnelle même si elle n'est pas *la plus* rationnelle. C'est, du moins, la théorie du conséquentialisme suffisant (*Satisficing Consequentialism*) de Michael Slote [Slote et Pettit, 1984 ; Slote, 1985].

Slote établit une analogie entre la moralité et la rationalité. Or, Herbert Simon (1916-2001) a montré que les individus qui se satisfaisaient de situations sous-optimales (des entreprises qui, par exemple, se contentaient d'un profit suffisant) ne contrevenaient pas nécessairement aux principes de la rationalité [Simon 1955]. La théorie du conséquentialisme suffisant prolonge donc cette thèse et conteste l'idée selon laquelle il faudrait que les individus fassent systématiquement le *maximum* pour aider les autres pour satisfaire aux conditions de la moralité. Même lorsque l'agent sait quelle option choisir pour maximiser l'utilité, il est parfaitement légitime qu'il en choisisse une autre qui lui paraît être suffisamment bonne. Michael Slote considère d'ailleurs que l'idée selon laquelle l'utilitarisme pourrait valoriser ce qui est mieux plutôt que ce qui est meilleur est déjà présente chez Bentham [Slote 1985].

Cette théorie a, cependant, été très critiquée [Hurka, 1990 ; Mulgan, 1993, 2001 ; Tucker, 2017]. L'analogie avec la théorie de la rationalité limitée, tout d'abord, est discutable. La thèse de Simon remet en cause l'un des postulats de l'économie néo-classique, celui selon lequel les agents sont parfaitement informés et qu'ils sont capables de prendre leurs décisions de façon parfaitement rationnelle. Il montre, qu'à défaut d'information parfaite, les agents mettent en place des heuristiques qui leur permettent de prendre de bonnes décisions, même si celles-ci ne sont pas optimales. Sa théorie n'a aucune dimension normative. Simon n'affirme nulle part que les agents ont raison de préférer une procédure sous-optimale à une procédure optimale.

La notion selon laquelle une action est 'suffisamment bonne' (*good enough*) [Slote, 1985, p.36] est, ensuite, trop vague. Ainsi que le souligne Thomas Hurka, elle peut renvoyer à deux schémas d'interprétation totalement différents [Hurka, 1990]. Selon le premier, on dira qu'une action est suffisamment bonne si ses conséquences ont atteint un niveau minimal critique au-delà duquel tout effort devient surrogatoire. Selon le second, on dira

qu'une action est suffisamment bonne si ses conséquences sont suffisamment proches du niveau optimal. Dans un cas, le jugement moral se détermine sur des niveaux absolus d'utilité, dans l'autre sur des niveaux relatifs. Les deux raisonnements sont possibles mais ils ont des implications morales très différentes, et peuvent difficilement être combinés.

Enfin, si de nombreux commentateurs s'accordent sur le fait que l'utilitarisme classique est trop exigeant, ils s'accordent aussi souvent sur le fait que le *satisficing*, lui, ne l'est pas assez. Le principe d'impartialité implique généralement que l'agent sacrifie son utilité pour les autres, ce qu'il a parfois de bonnes raisons de ne pas faire [Scheffler, 1982]. Se satisfaire d'une situation sous-optimale ne constitue, cependant, pas une raison valable pour se soustraire au principe d'impartialité [Mulgan, 2001]. Un agent ayant deux options, qui lui ont pour lui le même coût ou bénéfice, mais qui ont un impact fort différent sur le reste du monde, peut-il légitimement choisir l'option la moins bénéfique pour les autres au nom d'un principe de suffisance ?

Philip Pettit, quant à lui, a développé une théorie conséquentialiste qu'il définit comme une version restreinte (*restricted*) et infidèle du conséquentialisme (Pettit préfère le terme de conséquentialisme à celui d'utilitarisme, notamment à cause de sa critique de l'hédonisme). L'idée centrale de cette théorie est que si le processus de maximisation d'utilité standard permet d'évaluer et de comparer les différentes options qui s'offrent à l'agent, il ne lui permet néanmoins pas nécessairement de les sélectionner. Paradoxalement, la meilleure façon de satisfaire le critère de l'évaluation (ici la maximisation) peut être d'en restreindre son application [Pettit et Brennan, 1986, p.439]. Cet argument n'est pas entièrement nouveau, puisqu'on le retrouve notamment chez Sidgwick à propos de l'hédonisme, mais Pettit et Brennan en élargissent le champ d'application et lui donnent une importance accrue par rapport à leurs prédécesseurs. Il inclut notamment à l'évaluation des effets de l'action celle des prédispositions ou des motifs de l'agent. En plus d'identifier la meilleure action à mener, l'agent maximisateur doit également se demander si certaines de ses prédispositions ou traits de caractère ne contribuent pas également, dans la pratique, au processus de maximisation, et si celles-ci ne méritent pas d'être préservées ou développées.

Pourquoi l'agent maximisateur préférerait-il renoncer au calcul maximisateur pour se fier, par exemple, à ses prédispositions au risque d'agir de façon sous-optimale ? Parce que, selon Pettit et Brennan, le raisonnement maximisateur peut, sous certaines conditions, s'avérer contre-productif. C'est le cas lorsque les bénéfices d'une prédisposition sont hors d'atteinte par le calcul (*calculatively elusive*) mais également vulnérables au calcul (*calculatively vulnerable*). Certaines actions à forte utilité, produites par une prédisposition (à respecter ses engagements, par exemple) n'auraient simplement pas été possibles si elles avaient été envisagées par le calcul maximisateur. A cela s'ajoute le fait que l'idée même de choisir rationnellement d'agir selon cette prédisposition, plutôt que par le calcul, annule le bénéfice de l'opération. Il est donc relativement aisé de montrer que, dans un certain nombre de cas, il est préférable pour le conséquentialiste de restreindre son recours au calcul maximisateur. Or, selon Pettit, le respect des droits des personnes fait partie de ces prédispositions qui permettent de maximiser l'utilité totale. Selon lui, la prise en compte des droits des individus, droits qui – comme chez Dworkin – ont priorité sur l'utilité, peut s'avérer bénéfique au processus de maximisation.

### *Conclusion*

Pour conclure, nous avons tenté de présenter dans ce chapitre les principales questions liées à la mesure et à la maximisation de l'utilité. Nous avons pu observer que le principe *a priori* simple de Bentham « le plus grand bonheur pour le plus grand nombre » posait, en réalité, un grand nombre de problèmes techniques selon que l'on adopte une approche subjective ou objective de l'utilité, selon que l'on adopte une approche strictement hédoniste ou non, ou selon que l'on essaie de maximiser (ou non) l'utilité moyenne ou totale, ou encore selon l'horizon temporel sur lequel on maximise etc. Mais au-delà de ces questions strictement techniques, on est aussi en droit de s'interroger sur la nécessité même d'exiger d'une action qu'elle maximise l'utilité pour être considérée comme morale. Un certain nombre d'auteurs préfèrent d'ailleurs défendre une approche 'modérée' ou 'restreinte' de l'utilitarisme. Celles-ci rencontrent, cependant, d'importantes difficultés. Dans le chapitre suivant, nous présentons une approche alternative de l'utilitarisme, qui à l'instar du conséquentialisme modéré, tente de répondre aux problèmes liés à la

maximisation de l'utilité et aux objections morales dont fait l'objet l'utilitarisme. Il s'agit de l'utilitarisme de la règle.



### III. L'utilitarisme de la règle

L'une des plus importantes évolutions théoriques de l'utilitarisme a eu lieu vers la seconde moitié du vingtième siècle. Pensé autour d'une structure conséquentialiste directe, l'utilitarisme se réinvente en conséquentialisme indirect pour répondre aux différentes objections qui lui ont été adressées. Il existe de nombreuses versions de l'utilitarisme de la règle, chacune répondant à une conception particulière de ce que constitue une règle ou à une défaillance de l'utilitarisme de l'acte que l'auteur se propose de rectifier. Cet ensemble théorique est largement dominant aujourd'hui, mais il n'en est pas moins indemne de toute critique.

#### De l'utilitarisme de l'acte à l'utilitarisme de la règle

L'utilitarisme de la règle désigne un ensemble de théories qui ont en commun le caractère indirect et/ou collectif de l'évaluation morale des conséquences de l'action. Le développement de ces théories est assez bien circonscrit au niveau historique. Au niveau théorique, la variété des versions proposées dépend en grande partie de la conception de la règle utilisée par l'auteur.

#### *Qu'est-ce que l'utilitarisme de la règle ?*

L'idée d'un utilitarisme de la règle est développée, et défendue, pour la première fois par Roy Harrod dans son article 'Utilitarianism Revised' (1936). Mais celui-ci eu peu d'influence, et il faudra attendre la seconde moitié du XX<sup>e</sup> siècle pour que la question de la généralisation de l'utilitarisme devienne centrale, notamment avec les travaux de S.E. Toulmin, (*An Examination of the Place of Reason in Ethics*, 1950), M.G. Singer (*Generalization in Ethics*, 1961), David Lyons (*Forms and Limits of Utilitarianism*, 1965) et Richard Brandt (*Ethical Theory : The Problems of Normative and Critical Ethics*, 1959).

La distinction terminologique entre un utilitarisme de l'acte et un utilitarisme de la règle s'est officialisée à partir des années 1950 notamment sous l'impulsion de Richard Brandt en 1959. L'absence de distinction formelle entre ces deux types de conséquentialisme avant le vingtième siècle a conduit de nombreux commentateurs à s'interroger sur la position des fondateurs de l'utilitarisme envers l'un et l'autre. John Stuart Mill est ainsi soit présenté comme un tenant de l'utilitarisme de la règle [Urmson 1953], soit comme un tenant de l'utilitarisme de l'acte [Crisp 1997, Brink 2013] soit comme ne défendant ni l'un ni l'autre [Lyons 1994]. De la même façon, et bien qu'il soit présenté de façon prédominante comme un utilitariste de l'acte [de Lazari-Radek et Singer 2014] Sidgwick est parfois associé à l'utilitarisme indirect [Schultz 2004]. Bentham est, à quelques exceptions près, généralement considéré l'archétype de l'utilitarisme de l'acte.

L'utilitarisme de la règle, dans son acception générale, peut se définir de la façon suivante : « une action est bonne (*right*) si elle est conforme à une règle, qui, *si elle est suivie*, produit au moins autant d'utilité totale que n'importe quelle autre règle possible ». Il se distingue de l'utilitarisme de l'acte que l'on peut définir ainsi « une action est bonne (*right*) si elle produit au moins autant d'utilité totale que n'importe quelle autre action possible ». Dans ce dernier cas, le jugement moral sur l'action est déterminé par l'évaluation des conséquences de l'action en question sur le bien-être collectif (*utilitarisme de l'acte*). Dans l'autre, le jugement moral sur l'action est déterminé par l'évaluation des conséquences sur le bien-être général qui seraient produites par la règle à laquelle l'action en question appartient, si cette règle était *normalement* suivie (*utilitarisme de la règle*). Ainsi, si un mensonge, par exemple, peut se justifier par l'utilité nette qu'il peut produire (*utilitarisme de l'acte*), il ne peut pas se justifier moralement s'il est prouvé que la règle 'dire la vérité' produit au moins autant d'utilité nette que la règle 'ne pas dire la vérité'.

Il est important de distinguer plusieurs conceptions de la règle. Une règle peut, tout d'abord, désigner une *pratique* morale et/ou sociale. Ce type de règle n'a pas de fondement utilitariste explicite mais peut néanmoins contribuer *de facto* à maximiser l'utilité. Une règle peut, également, désigner une *heuristique*, c'est-à-dire un raccourci cognitif qui permet d'éviter un calcul de type conséquentialiste. Il ne s'agit donc pas nécessairement de règles morales. Ces règles, aussi appelées 'règles de pouce' (*rules of thumb*), jouent un

rôle prépondérant dans la prise de décision. Leur fonctionnement a beaucoup été étudié, notamment grâce aux travaux d'Herbert Simon et plus récemment d'Amos Tversky et de Daniel Kahneman. Dans la mesure où leur efficacité est attestée (même si elles ne sont pas nécessairement optimales), et qu'elles permettent aux individus d'éviter des stratégies plus coûteuses (en temps, en énergie, en information) et plus risquées (anticipations erronées etc), elles se justifient. Enfin, une règle peut désigner un *principe* normatif de portée universelle (ex. 'tu ne tueras point') ou de portée restreinte (ex. 'à moins que tu sois en situation de légitime défense, tu ne tueras point'). C'est généralement à cette troisième acception que l'utilitarisme de la règle, au moins dans ses formes primitives [Harrod, 1936], fait référence. Les formes plus contemporaines font plutôt référence aux pratiques [Hooker, 2000] ou à l'articulation entre règles théoriques et pratiques [Hare, 1981].

Les premières formes d'utilitarismes de la règle reposent sur une simple généralisation du raisonnement utilitariste classique (*cf* encadré 3). Au lieu de s'interroger sur l'impact d'une action particulière (i.e. mentir à son oncle Alfred sur l'état de sa santé), on s'interroge sur l'impact du mensonge en général. Pour David Lyons, cependant, cet utilitarisme de la généralisation constitue une forme 'primitive' de l'utilitarisme de la règle et ne devrait pas être confondu avec ce dernier [Lyons, 1965]. Pour Lyons, ce type d'utilitarisme ne se distingue pas fondamentalement de l'utilitarisme de l'acte.

Les formes contemporaines d'utilitarisme de la règle se rapportent plutôt à un ensemble de règles plutôt qu'à une règle en particulier. On parle alors de *code moral*. Ce code moral peut être *effectif*, c'est-à-dire existant et socialement accepté, ou *idéal*, autrement dit optimal au regard de l'utilité totale (*optimific*). Le code idéal est un ensemble de règles idéales qui sont acceptées ou internalisées par la très grande majorité de la population. Richard Brandt et Brad Hooker sont considérés comme les principaux théoriciens de l'utilitarisme du code idéal [Brandt, 1959 ; Hooker 2000]. Les théories du code *réel* (*actual*), moins répandues, ne prennent en compte que les règles déjà existantes [Miller, 2009 ; Toumin, 1950]. De ce fait, elles ne s'inscrivent pas dans une démarche maximisatrice mais appartiennent plutôt à l'utilitarisme modéré (*satisficing*).

Enfin, certains auteurs considèrent que le raisonnement conséquentialiste ne devrait pas se limiter à un ensemble de règles ou de pratiques, mais devrait également inclure les

dispositions ou les motivations des individus. On parle alors plutôt d'utilitarisme *global* [Driver,2014a]. Ainsi, selon David Brink, « [T]out objet d'une évaluation morale doit être évalué, non pas au regard de la valeur de ses conséquences pour le bonheur général, mais par sa conformité à quelque chose d'autre (i.e. des normes, des motifs ou des réactions) dont la valeur d'adhésion (*acceptance value*) est bonne ou optimale [Brink, 2010, p.1671]. L'utilitarisme global se veut beaucoup plus réaliste que l'utilitarisme de la règle *stricto sensu*. Il permet également de mieux prendre en compte la dimension collective de l'action éthique. Cet aspect est particulièrement important dans les débats sur la justice intergénérationnelle, où chaque individu ne peut prendre la pleine responsabilité morale du réchauffement climatique, par exemple. Nous aurons l'occasion de développer ce point dans le chapitre 4.

### Encadré 3 : Utilitarisme de la règle et généralisation utilitariste

Les premières formulations de l'utilitarisme de la règle se sont construites autour de la problématique de la généralisation : « Que se passerait-il si tout le monde faisait pareil ? ». Le questionnement est légitime puisqu'on a pu reprocher à l'utilitarisme de l'acte de justifier des actions, qui, si elles étaient généralisées, auraient des effets catastrophiques. C'est cette problématique qui a conduit Harrod en 1936 à proposer une nouvelle version de l'utilitarisme. Selon Harrod, le calcul maximisateur, pour être correct, devait penser l'utilité de l'action, non pas simplement en tant qu'utilité d'un acte isolé, mais en tant qu'utilité d'une instance d'un type d'action. L'utilité d'une action n'est, en effet, pas indépendante de l'utilité générale du type d'action auquel elle appartient. Autrement dit, une promesse n'a de valeur pour ceux qui la font/reçoivent que parce que la promesse, en tant que type d'action, a une forte utilité sociale. Inversement, les conséquences du non-respect d'une promesse ne peuvent être mesurées que vis-à-vis des conséquences de cette seule action mais doit également intégrer l'impact que cette violation peut avoir sur l'utilité sociale de la promesse en tant que type d'action.

Harrod propose d'abandonner la version 'grossière' (*crude*) de l'utilitarisme et d'y intégrer la formule kantienne de l'universalisation. Il définit le principe de sa version 'raffinée' (*refined*) de l'utilitarisme ainsi : « *Il existe certaines actions qui, lorsqu'elles sont réalisées en n occasions similaires ont des conséquences qui sont plus que n fois supérieures à celles produites lors d'une seule de ses occurrences* » [Harrod, 1936, p.48]. Une action particulière qui a une utilité nette positive mais qui, répétée *n* fois dans des circonstances similaires, aurait une utilité nette totale négative doit donc être considérée comme moralement 'mauvaise' (*wrong*). Cette thèse n'implique pas nécessairement, cependant, que les règles morales doivent être systématiquement obéies. Tout d'abord, il est théoriquement possible que l'infraction répétée d'une règle ait une utilité totale nette. Ensuite, et surtout, ce principe permet de distinguer un certain nombre de circonstances pour lesquelles il est

bon d'enfreindre la règle. On peut ainsi à la fois maintenir qu'il est bon de dire la vérité et en même temps considérer qu'il est possible (voire souhaitable) de mentir aux enfants ou aux personnes mourantes.

Cette conception de l'utilitarisme, également adoptée par Marcus Singer, a été très critiquée. Dès 1916, Charlie Dunbar Broad (plus connu sous le nom de C.D. Broad), conteste le raisonnement selon lequel on devrait intégrer dans le calcul de maximisation les conséquences d'actions collectives irréelles (et dont la plausibilité n'est pas même avérée). Broad appelle ceci le principe de fausse universalisation et dénonce l'usage des fausses hypothèses en éthique [Broad, 1916]. Selon Broad, l'esprit de la philosophie utilitariste est fondamentalement empirique et scientifique et il est illusoire de fonder ses calculs sur des scénarios fictionnels voire surréalistes. Il est alors absurde de se poser la question « que se passera-t-il si tout le monde fait la même chose ? » sachant que ce ne sera jamais le cas.

La seconde objection portée à la généralisation de l'utilitarisme a été portée par David Lyons en 1965 dans son ouvrage *Forms and Limits of Utilitarianism*. Selon Lyons, la généralisation de l'utilitarisme est une forme 'primitive' de l'utilitarisme de la règle et elle ne doit pas être confondue avec ce dernier. Elle ne constitue pas une rupture épistémologique avec l'utilitarisme de l'acte puisqu'elle est une simple extension logique. Si la maximisation d'utilité justifie le fait que je mente dans la situation x (par exemple à mon enfant pour Noël ou à mon oncle mourant), alors il existe nécessairement une règle selon laquelle il n'est pas permis de mentir sauf dans les toutes les situations de type x. Selon cette version de l'utilitarisme de la règle, il existe autant de règles que situations ou d'actions types. Derrière chaque règle, se profile en fait la maximisation d'utilité d'une action particulière mais généralisée. En d'autres termes, il n'y a pas de différences structurelles entre une généralisation utilitariste et l'utilitarisme de l'acte : les deux théories sont 'par extension équivalentes' (*extensionally equivalent*).

## **Principaux modèles de l'utilitarisme de la règle**

Nous avons vu dans la section précédente que l'utilitarisme de la règle recouvre un certain nombre de théories parfois très différentes les unes des autres. En voici trois instances : l'utilitarisme à deux niveaux de Hare, le code moral idéal de Brandt et Mulgan et, enfin, l'utilitarisme de la règle 'de type' contractualiste.

### *L'utilitarisme à deux niveaux de Hare*

Richard Hare (1919-2002) fut l'un des philosophes anglo-saxons les plus commentés dans les années 70 et 80. Il développe une philosophie morale, appelée prescriptivisme

universel, qui repose sur trois éléments : (1) les jugements moraux sont des prescriptions qui expriment des préférences, (2) les jugements moraux sont universalisables et (3) les jugements moraux l'emportent sur toutes autres préférences [Hare, 1981]. Autrement dit, toutes les préférences sont prescriptives mais seules celles qui sont universalisables et qui sont à même de l'emporter sur toutes les autres peuvent revendiquer le statut de jugements moraux. La démonstration de Hare est complexe et ambitieuse puisqu'elle tente de concilier une position réaliste ou objectiviste (au niveau méta-éthique) avec une théorie humienne de la motivation [Clementz, 2004].

Hare fait état du même esprit de conciliation au niveau de l'éthique normative. Son modèle, qui est à mi-chemin entre un utilitarisme de l'acte et un utilitarisme de la règle, présente deux niveaux. Le premier niveau, celui des 'archanges' (*archangels*), est *critique*, c'est-à-dire qu'il se fonde uniquement sur la délibération morale. C'est celui auquel les utilitaristes classiques ou les utilitaristes de l'acte font implicitement référence. Il implique une information parfaite et d'excellentes capacités cognitives. Comme ses prédécesseurs immédiats, Hare pense que les anticipations et les calculs nécessaires rendent ce processus de décision hors de portée au commun des mortels. Il reconnaît, à la suite de Sidgwick qu'il serait d'ailleurs contre-productif pour un homme de régler son comportement sur une arithmétique aussi complexe.

Le second niveau, celui des 'prolos' (*proles*) ou 'hommes du commun' (une référence à 1984 de George Orwell) est routinier et n'exige aucun calcul complexe. En pratique, nous suivons tous un certain nombre de règles préétablies, plus ou moins générales, qui nous permettent de traiter chaque nouvelle situation comme s'il s'agissait d'une situation connue et d'assurer *grosso modo* l'efficacité de nos actions. Ces règles, que Hare qualifie de *prima facie* ('à première vue', 'de prime abord'), nous sont intuitivement connues mais elles sont *en général* conformes aux prescriptions du niveau critique. Tout se passe *comme si* les règles que nous avons internalisées avaient été pensées par un esprit supérieur pour maximiser l'utilité. Lorsque la pratique n'est plus optimale, suite à une évolution du contexte, des mœurs etc., celle-ci est révisée au niveau critique, puis internalisée à nouveau. Cela permet aux agents de maximiser l'utilité de leurs actions sans avoir à délibérer sur la valeur de leur action particulière.

En réalité, les êtres humains sont tous mi-archanges et mi-prolos. Leur processus de décisions possède à la fois un niveau critique et un niveau intuitif. La réflexion critique, conforme au calcul utilitariste standard, est, en revanche, préférable lorsque nous disposons des facultés cognitives, des informations et du temps nécessaires à la délibération. Lorsque ces conditions ne sont pas remplies et/ou lorsque la situation nous est bien connue (action routinière), il est préférable de suivre notre intuition, c'est-à-dire les règles que nous avons internalisées. En cas de conflit entre plusieurs règles 'prima facie', le recours au niveau critique est nécessaire pour arbitrage. Notons que le niveau critique se forme avec l'âge et l'éducation. C'est la raison pour laquelle on inculque aux enfants des règles strictes (par exemple ne traverser la rue que lorsque le signe piéton est vert) et que nous y dérogeons lorsque cela nous paraît utile et justifié, (par exemple, quand il n'y aucune voiture en vue).

Selon Hare, nous sommes en permanence en train de basculer d'un niveau à l'autre. Les deux niveaux sont parfaitement complémentaires. Parfois nous agissons comme des utilitaristes de l'acte, parfois comme des utilitaristes de la règle. Bien que sa théorie légitime l'usage des règles, celles-ci restent tributaires du niveau critique.

### *Le code moral idéal*

Ainsi qu'on l'a vu dans la section précédente, on distingue deux types d'utilitarisme de la règle selon qu'ils renvoient à un code moral idéal (*ideal moral code*) ou un code moral réel (*actual moral code*). Les théories du code moral idéal sont dominantes. Richard Brandt est le premier à l'avoir développé [Brandt, 1979, 1992]. Il est aujourd'hui défendu par Brad Hooker et Tim Mulgan [Hooker, 1990, 2000 ; Mulgan, 2001 ; 2020].

L'utilitarisme de la règle idéale se distingue à la fois de l'utilitarisme de la règle primitif, au sens de Lyons (autrement dit de l'utilitarisme généralisé), et de l'utilitarisme du code moral réel. A l'instar de Lyons, ses partisans considèrent qu'une simple généralisation (au sens kantien du terme) du raisonnement utilitariste classique (celui de l'acte), ne présente aucun avantage par rapport à l'utilitarisme de l'acte, auquel elle se réduit *in fine*. Les théories de l'utilitarisme du code moral réel, défendues notamment par Toulmin, Rawls,

Strawson ou Urmson, présentent également de nombreux inconvénients, dont celui de justifier, dans certains contextes socio-historiques, des règles profondément immorales. Ils opposent aux codes moraux réels un code moral idéal dont l'adoption générale maximiserait l'utilité, même si – en pratique – ce code n'est pas suivi par tous ni dans toutes les situations données. Brandt et Hooker considèrent ainsi qu'idéalement le code moral doit être accepté par au moins 90% de la population.

Quelles sont les caractéristiques d'un code moral idéal ? Tout d'abord, c'est un code qui est suffisamment simple pour être adopté et internalisé par les agents. Bien qu'il puisse occasionnellement s'accorder avec les règles institutionnelles de la société, il en est cependant fondamentalement distinct et peut ainsi parfois permettre de les dénoncer. L'utilité de ce code ne se mesure ensuite pas simplement à l'aune du gain d'utilité totale qu'il peut produire mais également aux coûts qu'il peut générer chez les individus, y compris les coûts liés à son enseignement. Pour être efficace, un code moral doit, en effet, être internalisé. Les théoriciens du code moral idéal distinguent les règles qui sont *suivies* des règles qui sont *acceptées*. On peut obéir à une règle sans l'accepter. Inversement, on peut l'accepter sans toujours la suivre. Mais, pour qu'un code moral soit réellement efficace, il doit être accepté ou internalisé. L'internalisation des règles fait reposer le mécanisme d'optimisation sur les ressorts psychologiques impliqués dans la motivation morale, et plus spécifiquement le sentiment de culpabilité et le regret. En outre, cela garantit sa transmission entre les générations.

Accepter un code moral idéal est bon non seulement parce que les règles suivies maximisent l'utilité totale mais également parce qu'en internalisant la règle on agit indirectement sur le comportement moral des autres. Brandt qualifie d'ailleurs sa théorie d'optimifque indirecte. Il compare sa version de l'utilitarisme du code idéal à l'utilitarisme de Mill, qui intègre, lui aussi, la sanction interne de la conscience (que Bentham avait rejetée), ainsi que les coûts en termes d'utilité générés par la réprobation sociale (ou, au contraire, la motivation induite par le désir d'être estimé).

Le code moral idéal est plus souple que la règle généralisée. Il est possible d'accepter un code sans en suivre systématiquement toutes les règles. Un code moral idéal, et donc optimal, peut 'tolérer' qu'une minorité agisse de façon immorale. Il tolère également que



les membres bien intégrés de la société s’y soustraient occasionnellement. Le code moral idéal n’est donc pas trop exigeant moralement. Il est aussi plus réaliste que la généralisation utilitariste. Mais cette grande flexibilité peut jouer en sa défaveur. Moins les individus le respectent, plus l’effort que doivent consentir les autres est important. Ce problème est particulièrement sensible en éthique intergénérationnelle [Mulgan, 2006].

### *Utilitarisme de la règle et contractualisme*

En plus des théories du code moral idéal et du code effectif, ainsi que des théories à deux niveaux qui associent utilitarisme de l’acte et de la règle, il existe, enfin, un troisième type de théories d’utilitarisme de la règle, théories qu’on peut qualifier de contractualistes même si elles se démarquent sensiblement du contractualisme rawlsien. Les plus importantes sont celles développées par John Harsanyi, Tim Scanlon et Derek Parfit. Leur fondement contractualiste peut parfois faire douter de leur nature utilitariste, sauf chez Harsanyi qui revendique fortement cette appartenance.

Harsanyi est le premier à associer utilitarisme et raisonnement contractualiste. Par raisonnement contractualiste, on entend un raisonnement qui s’appuie sur le consentement unanime d’agents impartiaux. L’impartialité stricte implique soit l’intervention d’un agent extérieur et non-intéressé (le spectateur impartial) soit le recours à une fiction du type voile d’ignorance. Bien avant Rawls, et sans employer cette terminologie particulière, Harsanyi développe un modèle dans lequel les agents sont amenés à se prononcer sur la désirabilité des principes de justice d’une société dans laquelle ils ne connaissent pas leur position future [Harsanyi, 1982]. A l’instar de Rawls, Harsanyi fait le pari que tous les agents choisiront rationnellement la même solution. Le parallèle s’arrête cependant là. Harsanyi critique sévèrement la conclusion de Rawls, qui affirme qu’en l’absence d’informations personnelles, les agents choisiront nécessairement de maximiser le sort des plus mal lotis, principe aussi appelé le *maximin*. De fait, il est fort possible que de nombreux agents choisissent une société fortement inégalitaire s’ils estiment que l’utilité espérée (probabilisée) d’y être gagnant est supérieure à celle d’y être perdant. Le risque que les agents sont prêts à prendre révèle la désirabilité du gain potentiel pour l’agent.

Dans l'approche d'Harsanyi, les préférences des agents sont connues (il s'agit d'une version de l'utilitarisme des préférences, cf chapitre 2). Il s'agit néanmoins de leurs préférences réelles ou informées, autrement dit les préférences que les agents auraient s'ils disposaient de toute l'information nécessaire, s'ils délibéraient avec la plus grande attention et dans un état d'esprit favorable au choix rationnel [Harsanyi, 1982, p.55]. Ces préférences excluent notamment les préférences anti-sociales telles que l'envie ou la jalousie. Si les préférences des agents sont connues, en revanche, leur position socio-économique dans la société envisagée ne l'est pas mais Harsanyi suppose que chacun a une même probabilité d'occuper une position donnée. On parle alors de modèle d'équiprobabilité des jugements de valeurs morales. Harsanyi montre alors que, selon ce modèle, l'ensemble des règles qui sera choisi par les agents sera celui qui produit le plus haut niveau d'utilité moyenne pour les membres de la société. Ce modèle permet ainsi de légitimer le système politique (démocratique), économique (capitaliste) mais également moral (utilitarisme de la règle) de nos sociétés.

Selon Harsanyi, l'utilitarisme de la règle est le système moral qui permet, y compris si tous les agents ne s'y conforment pas, de maximiser l'utilité sociale. Son principal avantage par rapport à l'utilitarisme de l'acte est de permettre une meilleure coordination entre agents. Il est également beaucoup plus flexible que la déontologie kantienne ou l'utilitarisme généralisé dans la mesure où il permet de justifier un certain nombre d'exceptions à la règle, et ainsi d'échapper à l'objection de fétichisme de la règle.

Dans son *Anthologie historique et critique de l'utilitarisme*, Catherine Audard accorde une place importante à la philosophie contractualiste de Thomas Scanlon qu'elle considère être une « variante de l'utilitarisme » [Audard, 1999, vol.3, p.96]. Cela se discute mais il n'en a pas moins joué un rôle essentiel dans son histoire récente, et c'est la raison pour laquelle nous avons choisi de l'intégrer ici. Scanlon rejette l'utilitarisme philosophique, autrement dit la thèse selon laquelle les seuls faits moraux fondamentaux sont ceux concernant le bien-être individuel [Scanlon, 1982, p.108]. Il lui oppose le contractualisme, selon lequel les faits moraux fondamentaux font l'objet d'un accord général, libre et informé. En soi le contractualisme ne réfute donc pas l'utilitarisme normatif puisqu'il est théoriquement possible que les règles établies selon les principes contractualistes soient tout à fait conformes aux prescriptions utilitaristes. Mais quelles que soient les implications

normatives du contractualisme, celles-ci ne sont pas motivées par des principes utilitaristes. Le bien-être total (ou moyen) ne constitue pas, dans l'approche contractualiste, un critère de justification de l'action.

Le raisonnement contractualiste et l'utilitarisme de la règle ont pu être comparés dans la mesure où l'un et l'autre s'intéressent aux conditions d'universalisation de l'action individuelle. Rawls et Harsanyi, par exemple, adoptent une posture similaire lorsqu'ils font de l'utilité espérée d'agents en situation d'incertitude ou d'ignorance le fondement de la rationalité et, par extension, de la morale. Mais, contrairement à Rawls, Scanlon ne considère pas la satisfaction de l'intérêt individuel comme un fondement adapté à une philosophie morale. Il ne cherche pas non plus à placer les agents dans des conditions telles qu'ils pourraient se mettre à la place des autres. Ils sont à leur place, et les raisons qu'ils ont de refuser une action dont ils pourraient être victimes – même si celle-ci a pour effet d'accroître l'utilité moyenne ou totale – n'en ont pas moins de légitimité. Notons, enfin, un dernier argument qui pourrait accréditer la thèse selon laquelle le contractualisme scanlonien est une forme de l'utilitarisme (de la règle), à savoir l'interprétation qu'en a fait Derek Parfit dans *On What Matters* [Parfit, 2011 ; Parfit, 2017].

Derek Parfit (1942-2017) développe l'idée selon laquelle les différentes théories que sont la déontologie kantienne, le contractualisme (scanlonien) et l'utilitarisme de la règle ne sont, en fait, que trois approches différentes d'une même 'réalité' morale, à l'image de ce que peuvent être les trois faces d'une même montagne, et que l'on a donc tort de les opposer l'une à l'autre. Notons immédiatement que Parfit n'emploie pas l'expression utilitarisme de la règle à laquelle il préfère celle de conséquentialisme de la règle. Contrairement aux philosophies déontologiques et contractualistes, qu'il associe respectivement à Kant et Scanlon, il est impossible d'identifier avec certitude les fondements théoriques du conséquentialisme de la règle qu'il semble défendre.

Parfit présente, tout d'abord, une version 'kantienne' du conséquentialisme de la règle qu'il définit ainsi : « Chacun devrait suivre les principes optimifiques, car ce sont les seuls principes que chacun pourrait rationnellement vouloir qu'ils deviennent des lois universelles » [Parfit, 2011, vol.1, p.413]. Il montre, ensuite (*Argument de la convergence*), que les principes que personne ne pourrait raisonnablement rejeter sont les mêmes que

ceux que tout le monde pourrait rationnellement vouloir [Parfit, 2011, vol.1, p.412], et conclut logiquement que ces trois ensembles théoriques peuvent se reformuler sous une seule thèse, qu'il nomme Triple Théorie, et qui établit qu'une action est mauvaise quand elle appartient à un type d'action qui n'est pas permis par quelque principe optimifique, universellement souhaitable par chacun (*uniquely universally willable*), et qui ne peut être raisonnablement rejeté par personne [Parfit, 2011, vol.1, p.413]. Ce que Parfit nomme Triple Théorie constitue une formulation de l'utilitarisme de la règle, ou plus exactement de l'utilitarisme de la règle dans sa version primitive (au sens de Lyons), à savoir l'utilitarisme de l'acte généralisé. Cela peut étonner puisque, on l'a vu, cette interprétation a été beaucoup critiquée dans le passé, et que la plupart des auteurs utilitaristes récents de Harsanyi à Brandt en passant par Hare ont tenté, chacun à leur façon, de montrer que l'utilitarisme de la règle ne se réduisait pas à un utilitarisme de l'acte. Les raisons qui ont poussé Parfit à revenir à cette version primitive de l'utilitarisme de la règle sont essentiellement d'ordre méta-éthiques [Salvat, 2019b]. Mais ce retournement est peut-être également la marque d'une évolution plus profonde de la pensée utilitariste contemporaine, qui à l'instar de Peter Singer, assume de plus en plus la responsabilité morale *directe* qu'ont les agents lorsqu'ils agissent ou s'abstiennent d'agir.

### **Avantages et limites de l'utilitarisme de la règle**

L'utilitarisme de la règle est, on l'a vu, multiforme. Il présente de nombreux avantages par rapport à l'utilitarisme de l'acte. Le premier d'entre eux est de garantir le respect d'un certain nombre de règles morales de base, comme respecter sa promesse, y compris si, dans un cas précis, une violation de cette règle serait optimale. L'utilitarisme de la règle présente également un second avantage certain sur l'utilitarisme de l'acte : être moralement moins exigeant que ce dernier.

#### *L'utilitarisme de la règle face aux objections faites à l'utilitarisme classique*

La première objection opposée à l'utilitarisme a été déjà évoquée dans le chapitre 2. La philosophie utilitariste peut être aliénante ou moralement trop exigeante. Les agents

peuvent, en effet, parfois être conduits à sacrifier leur utilité personnelle ou celle de leurs proches pour maximiser l'utilité totale. Le caractère excessif de l'exigence morale résulte, en partie, de l'objectif de maximisation imposé par l'utilitarisme classique. C'est la raison pour laquelle ont été développées des théories modérées ou restreintes de l'utilitarisme. Mais, ainsi qu'on l'a vu dans le chapitre précédent, il est difficile d'établir un seuil critique entre ce qui est suffisant et ce qui ne l'est pas, si bien que ces théories peuvent être accusées à la fois d'être toujours trop exigeante ou de ne plus l'être assez.

Par ailleurs, quel que soit le niveau d'utilité retenu, le conséquentialisme modéré, à l'instar de l'utilitarisme classique, continue à faire reposer sur l'agent la pleine responsabilité morale de ce qui adviendra dans le futur. Rappelons que, selon l'utilitarisme classique, les agents sont responsables des conséquences de leurs actions mais également de celles qui auraient pu être évitées par leurs actions. Dans les années 1970, Peter Singer avait ainsi défendu le principe de la responsabilité *négative* des individus, c'est-à-dire celle causée par une absence d'actions [Pariente-Butterlin, 2016 ; Singer, 1972]. Dans la perspective de ce que Tim Mulgan appelle un 'monde abîmé' (*broken world*), cette responsabilité individuelle est écrasante [Mulgan, 2011]. L'utilitarisme de la règle, aussi appelé utilitarisme indirect ou collectif, permet en partie de répondre à cette objection en faisant de la moralité un projet collectif, auquel chacun prend sa part en respectant les règles établies [Mulgan, 2020]. De façon plus précise, Brad Hooker, l'un des principaux tenants de l'utilitarisme de la règle aujourd'hui, considère que la pauvreté mondiale pourrait être éliminée si chaque personne aisée du monde développé donnait 10% de son revenu à des organisations caritatives [Hooker, 1990, p.72]. Lorsque la règle s'applique *collectivement*, il est possible de répartir le coût en termes d'utilité entre toutes les personnes.

La deuxième objection soulevée par les opposants de l'utilitarisme classique est son 'immoralité', ou plus exactement, le caractère moralement choquant de certaines de ses prescriptions. Ainsi que Sidgwick l'a montré, l'utilitarisme, dans une grande majorité de cas, confirme par la raison les prescriptions que nous imposent d'autres sources de la morale telles que la religion ou l'intuition. Dans un certain nombre de cas, cependant, il est possible de montrer que le raisonnement conséquentialiste nous amène logiquement à justifier des actions autrement considérées comme moralement inacceptables. Pour les utilitaristes classiques, il n'existe aucune règle morale autre que celle de la maximisation d'utilité. Si

celle-ci exige la violation d'une promesse, un mensonge ou, pire, le meurtre d'un innocent, alors ces actions traditionnellement considérées comme immorales sont légitimes. Or, pour certains philosophes, les interdits moraux sont indispensables à la vie en société [Anscombe, 1958].

Le fait qu'une règle soit adoptée sur un principe utilitariste ne permet cependant pas ensuite aux individus de juger de l'opportunité ou non de suivre cette règle. John Rawls, dans un article de 1955, sauve l'utilitarisme de l'objection selon laquelle l'utilitarisme justifie l'injustifiable, à savoir la condamnation d'un innocent [Rawls, 1955]. Rawls montre, en effet, que les critiques de l'utilitarisme (comme parfois les utilitaristes eux-mêmes) n'attachent pas suffisamment d'importance à la différence entre justifier une règle et justifier les actions qui relèvent de celle-ci. Le principe utilitariste de maximisation de l'utilité totale peut, en effet, justifier une règle (par exemple respecter ses promesses) sans justifier les actions individuelles qui relèvent du respect (ou non) d'une promesse. Le fait qu'une pratique (respecter une promesse) soit justifiée par un principe utilitariste ne donne absolument pas à l'agent qui s'y adonne (qui fait une promesse) la liberté d'utiliser des arguments utilitaristes pour décider s'il va ou non honorer sa promesse.

La dernière objection faite à l'utilitarisme classique est connue sous le nom d'argument de Hodgson. L'argument est potentiellement destructeur car il affirme que la théorie utilitariste n'est pas généralisable et qu'elle ne peut donc prétendre au titre de théorie éthique. Il est développé en 1967 par D.H. Hodgson dans *Consequences of Utilitarianism* [Hodgson, 1967]. Son argument repose sur les effets délétères qu'une adhésion généralisée à la théorie utilitariste classique, plus précisément un utilitarisme de l'acte, peut avoir sur les conséquences des actions des agents. Hodgson prend pour exemple la convention morale qui condamne le mensonge et la rupture de promesse. Faire une promesse n'a de valeur que si la personne à qui on promet croit que celle-ci sera réalisée quelles qu'en soient les conditions. Dès lors qu'il est de connaissance commune que seules les actions qui maximisent l'utilité seront réalisées, le fait même de promettre quelque chose n'a plus de sens. De la même façon, dire la vérité n'a d'intérêt que si vos interlocuteurs croient que vous dites la vérité. Si les individus estiment que la véracité des informations données dépend de son impact potentiel sur l'utilité totale, ils n'auront plus de raison de s'y fier. Cela conduit au paradoxe suivant, qui n'est pas sans rappeler le

paradoxe de l'hédonisme de Sidgwick : une société qui respecte les règles de morales conventionnelles est plus à même de maximiser l'utilité totale qu'une société dans laquelle tous les habitants sont des utilitaristes de l'acte. Une bonne théorie utilitariste serait donc une théorie capable de justifier l'existence de règles morales de type déontologique, autrement dit dont le respect est inconditionnel, telles que 'il faut toujours dire la vérité' ou 'il faut toujours respecter ses promesses'.

### *Limites de l'utilitarisme de la règle*

Malgré ces atouts, l'utilitarisme de la règle fait aussi l'objet de sévères critiques. Selon certains partisans de l'utilitarisme de l'acte, dont John J.C. Smart, obéir à des règles construites dans le but de maximiser l'utilité, y compris lorsque cela est désavantageux pour tous, relève du fétichisme de la règle. Pour Smart, les règles morales sont utiles lorsqu'il s'agit d'heuristiques, c'est-à-dire dès lors qu'elles permettent aux individus de prendre des décisions rapides et/ou lorsque l'information disponible est insuffisante. Elles sont avantageuses dans la mesure où leur prescription a une probabilité raisonnablement élevée d'être conforme à celle qu'aurait produite une délibération rationnelle. Mais, dès lors qu'il est avéré que les prescriptions de la règle sont contraires à la maximisation d'utilité, il devient – selon Smart – irrationnel de suivre la règle : « N'est-ce pas ériger R [la Règle] en une sorte d'idole que de continuer à s'y conformer lorsque s'y soustraire pourrait empêcher, disons, une souffrance évitable ? Cela ne relève-t-il pas d'une forme superstitieuse de fétichisme de la règle (facilement explicable psychologiquement) et non pas de la pensée rationnelle d'un philosophe ? » [Smart, 1956, p.348-49]. Selon Smart, dans la mesure où il implique une obéissance absolue – et parfois irrationnelle – à la règle, l'utilitarisme de la règle tient davantage de la déontologie que du conséquentialisme.

On peut, aussi, lui reprocher de trop concéder aux thèses intuitionnistes. Les théories du code moral, en particulier, ont pour particularité d'approuver un ensemble de règles, existantes ou non, pas nécessairement optimales *stricto sensu*, mais qui ont pour avantage d'être facilement acceptées et enseignées. Ces règles sont, en pratique, celles que nous avons déjà plus ou moins intériorisées. Une règle dont le résultat serait très contre-intuitif (comme, par exemple, le fait de préférer un étranger à son enfant) aurait, en effet, un coût

d'apprentissage beaucoup trop élevé. Si l'utilitarisme de la règle réconcilie l'utilitarisme avec l'intuitionnisme, ce que Sidgwick considérait comme un élément positif, on peut néanmoins lui objecter le fait que nos intuitions peuvent également être profondément injustes ou arbitraires. De ce point de vue, l'utilitarisme de la règle est en porte-à-faux avec l'esprit initial de l'utilitarisme tel qu'il était porté par Bentham. Le fait que les normes intuitives peuvent parfois indirectement contribuer à la maximisation de l'utilité (Sidgwick) ne doit pas être confondu avec le fait que les règles morales doivent se conformer à un code moral intuitivement validé pour optimiser son observance.

#### *Au-delà de l'utilitarisme de la règle*

L'utilitarisme de la règle tente de réconcilier la doctrine utilitariste avec la morale du sens commun, d'une part, et avec la déontologie kantienne, de l'autre. L'utilitarisme du code moral est, de toutes les versions, celle qui a probablement le mieux réussi cette transformation. Mais, ainsi que nous venons de le voir, celle-ci n'est pas sans provoquer des remous parmi les utilitaristes. Ne serait-on pas aller trop loin du côté de la déontologie ? Aurait-on laissé de côté l'essence rationaliste de l'utilitarisme classique ? C'est le sens de la critique de Smart.

Mais au-delà de cette question très théorique s'en pose une autre, plus pratique, aujourd'hui. L'utilitarisme de la règle a souhaité montrer que l'utilitarisme n'impliquait pas forcément des sacrifices excessifs de la part des individus. Les théories du code moral idéal tolèrent même qu'une partie de la population y déroge totalement. Mais n'est-ce pas tomber là dans l'excès inverse ? Cela ne revient-il pas à déresponsabiliser les individus ? La question est posée par un certain nombre de théoriciens et de militants qui dénoncent l'attitude morale qui consiste à ne pas suffisamment se poser la question de notre responsabilité individuelle et à se défausser sur la responsabilité collective. Le renouvellement de l'éthique appliquée, centrée aujourd'hui sur les questions de bioéthique, de souffrance animale et de réchauffement climatique (cf chapitre 4), a probablement beaucoup contribué à cette réévaluation de la responsabilité morale. Elle ne se traduit pas (encore) par un regain de l'utilitarisme de l'acte, mais certains développements s'orientent dans cette direction.



Le premier d'entre eux est, peut-être, le pessimisme affiché par certains théoriciens de l'utilitarisme de la règle, dont Tim Mulgan, qui vont jusqu'à admettre que le principe d'obéissance partielle (*partial compliance*) ne permet pas à l'utilitarisme de la règle de se constituer une théorie morale complète [Mulgan, 2006, p.302]. Mulgan propose de diviser la morale entre deux branches, le domaine de la nécessité et celui de la réciprocité [Mulgan, 2001]. Les situations de famine, par exemple, relèvent du domaine de la nécessité. Les situations de la vie ordinaire relèvent, quant à elles, plutôt du domaine de la réciprocité. Mulgan admet que l'utilitarisme de la règle ne permet pas de répondre efficacement à toutes les questions éthiques, et que, dans certains cas, les individus doivent faire le maximum possible pour aider les autres.

Le rapprochement de l'utilitarisme de la règle et du contractualisme pourrait également être le signe d'une évolution, sinon d'un affaiblissement, de l'utilitarisme de la règle. Le contractualisme est, en effet, plus contraignant que le code moral. Les règles sont pensées et acceptées par l'individu, et ce dernier a obligation de les respecter. La règle d'universalisation kantienne, intégrée par Parfit, a par ailleurs pour conséquence de construire chaque règle à partir de la généralisation d'une action. Or, ainsi que Lyons l'a montré, la généralisation utilitariste a une structure logique identique à celle de l'utilitarisme de l'acte. Le retour de ce que Lyons a qualifié de forme 'primitive' de l'utilitarisme de la règle peut surprendre. Mais, malgré ses nombreuses limites, la généralisation utilitariste a pour avantage de ne pas diluer la responsabilité morale des agents dans une pratique collective.

Enfin, un troisième développement théorique récent pourrait être interprété comme un signe d'affaiblissement de l'utilitarisme de la règle. Il s'agit de la théorie de l'altruisme efficace (*effective altruism*). L'altruisme efficace est un mouvement né dans les années 2000 sous l'impulsion conjointe de philosophes utilitaristes, tels que Peter Singer, William MacAskill ou Toby Ord, et de militants du milieu associatif [MacAskill, 2015 ; Singer, 2015]. La thèse de ce mouvement est que tout doit être fait pour éviter au maximum les souffrances humaines (mais aussi animales), et que les citoyens des pays riches, qui sont à la fois ultra-privilegiés et, du fait de leur style de vie et de leur comportement consumériste, en partie responsables des souffrances dans le monde, ont non seulement une obligation morale de vivre plus modestement, de donner une partie de leurs revenus

(10%) à des organisations caritatives, mais aussi de maximiser l'utilité de ce don, en sélectionnant (ou en suivant les recommandations de sites spécialisés) les organisations les plus à même d'utiliser ces dons efficacement.

### *Conclusion*

Nous avons exposé dans ce chapitre l'une des principales évolutions de l'utilitarisme contemporain par rapport à l'utilitarisme classique. L'utilitarisme de la règle présente de nombreux avantages par rapport à l'utilitarisme de l'acte. Il permet, tout d'abord, de justifier l'existence de normes et de pratiques morales conventionnelles sans lesquelles la vie en société deviendrait rapidement impossible. Il montre que l'utilitarisme, dans certaines conditions, peut être compatible avec la morale du sens commun, d'un côté, et avec la déontologie, de l'autre. L'utilitarisme de la règle permet également d'échapper à la critique selon laquelle les sacrifices que cette théorie éthique exige des individus sont excessifs. En intégrant dans le calcul utilitariste le fait que tous les autres individus participent également à l'effort nécessaire, la règle rend le projet moral collectif, plus raisonnable et par conséquent plus réaliste.

Malgré ces avancées, l'utilitarisme de la règle n'est pas totalement satisfaisant. Les questions relatives à la construction des règles, de leur acceptation et de leurs révisions restent souvent à clarifier. Une trop grande tolérance à l'égard de la morale du sens commun ou de la déontologie peut remettre en cause la nature conséquentialiste, rationnelle ou optimisatrice de cette théorie. Un trop grand laxisme à l'égard du respect des règles par les individus a pour conséquence d'alourdir les exigences et les sacrifices déjà consentis par ceux qui se plient de bonne volonté au code moral. Dans les faits, les limites de l'utilitarisme de la règle se révèlent être celles du code moral existant, il est injuste et sous-optimal.

Dans le chapitre qui suit, nous discutons du quatrième pilier de l'utilitarisme classique (après le maximalisme, le welfarisme et le conséquentialisme direct), à savoir l'impartialité, et nous nous poserons la question de savoir si les limites de l'utilitarisme ne sont pas, *in fine*, des corollaires de son principe d'impartialité.



## IV/ IMPARTIALITE: THEORIE ET PRATIQUE

Le quatrième pilier de l'utilitarisme classique est le principe d'impartialité. Celui-ci est étroitement lié à celui de l'égalité même si l'un n'implique pas forcément l'autre. La question de l'égalité, et notamment de l'égalité politique occupe une place prédominante dans la philosophie de Bentham, de Mill et de Sidgwick. La dimension politique ou sociale de l'utilitarisme s'est très largement érodée aujourd'hui au profit de sa dimension éthique. Celle-ci a d'ailleurs été profondément renouvelée ces dernières années par les débats sur la bioéthique, la justice intergénérationnelle et l'éthique animale.

En dépit de son positionnement théorique, l'utilitarisme a été très critiqué pour ses manquements envers le principe d'égalité. Cela a déjà été montré dans le cas de l'égalité politique chez Mill et Sidgwick notamment. En matière éthique, il leur a été objecté que, loin de contribuer à une répartition plus égalitaire des utilités, la maximisation contribuait parfois à aggraver les inégalités. Il leur a été également reproché qu'en niant l'existence des droits des individus, l'utilitarisme ne pouvait raisonnablement prétendre respecter leur égalité.

L'objet de ce chapitre est double. Il s'agit, tout d'abord, de présenter les fondements, les implications et les limites théoriques du principe d'impartialité dans la philosophie utilitariste. Il s'agit, ensuite, d'en illustrer l'importance à travers trois questions fondamentales de l'éthique appliquée : l'égalité entre les hommes (la pauvreté, les famines), l'égalité entre les générations d'hommes (le réchauffement climatique) et l'égalité entre les hommes et les animaux (la souffrance animale).

### **Le principe d'impartialité**

*Impartialité, universalité et égalité*

Selon le principe d'impartialité, une règle morale ne peut favoriser les intérêts de certaines personnes ou de certains groupes de personnes au détriment des autres. On définit alors l'impartialité comme *l'impossibilité de justifier un choix par les intérêts d'une ou de plusieurs personnes*. Nous aurons l'occasion ci-après de le distinguer de l'impersonnalisme que l'on définit comme *l'impossibilité de justifier un choix du point de vue d'une ou de plusieurs personnes*.

La notion d'impartialité est étroitement liée à deux autres notions, également très importantes dans le raisonnement utilitariste, celle de l'universalité et celle de l'égalité. Dans un certain nombre de cas, elles ne sont pas vraiment distinguées. Elles peuvent, pourtant, recouvrir différentes réalités. Le principe d'universalité, que l'on trouve déjà chez Kant, affirme que, pour qu'elle puisse être qualifiée de morale, une action doit pouvoir être universalisable, c'est-à-dire faire l'objet d'une loi universelle, s'appliquant à tous et dans toutes les circonstances. En éthique, on définit donc l'universalité à *l'impossibilité de justifier un choix à partir de situations particulières*.

A part quelques rares philosophes, dont Parfit, les utilitaristes n'ont pas recours à la notion d'universalité. Les utilitaristes de la règle lui préfèrent celle de généralité. On dit qu'une action peut être qualifiée de morale si (en plus des critères de maximisation qui la caractérisent), celle-ci est généralisable, c'est-à-dire qu'elle n'est particulière ni à une situation ni à une personne donnée. Elle peut, néanmoins, être particulière à un type de situations ou de personnes. A la différence de l'universalisation, le processus de généralisation ne garantit pas l'impartialité des choix.

En pratique, l'impartialité consiste à évaluer les conséquences d'une action 'du point de vue de l'univers', selon l'expression de Sidgwick, ou encore, à se mettre à la place de toutes les personnes potentiellement affectées par ladite action. Autrement dit, soit on ne prend parti pour aucune personne, soit on prend parti pour toutes à la fois. Il existe donc un lien logique entre impartialité et égalité. Peter Singer considère ainsi que l'adoption de ce principe le conduit directement à adopter celui d'égalité de considération des intérêts individuels [Singer, 1979]. Nous verrons, néanmoins ci-dessous, que le fait de ne pas vouloir prendre parti ne signifie pas nécessairement traiter tous les individus de façon égale.

### *Impartialité et maximisation d'utilité*

L'impartialité n'est pas spécifique à l'utilitarisme. Un certain nombre de philosophes pensent d'ailleurs qu'elle est indissociable de tout raisonnement moral [Singer, 1993, p.11]. Le contractualisme est typiquement une théorie fondée sur l'impartialité puisque les règles morales sont consenties par des individus alors qu'ils sont dépourvus de toute caractéristique personnelle, sociale ou pouvant être socialement pertinentes (sexe, ethnie, éducation, richesse etc.). C'est ce que l'on appelle le voile d'ignorance.

Si le principe d'impartialité n'est pas spécifique à l'utilitarisme, en revanche, la manière dont il se manifeste dans ses prescriptions l'est, et a été l'objet de fortes controverses. Associé au principe de maximisation d'utilité, l'impartialité implique, en effet, le fait de ne pas prendre en compte l'identité des agents impactés mais uniquement la quantité d'utilité produite (ou détruite) par une action. William Godwin (1756-1836), l'un des premiers représentants de l'utilitarisme, a donné un exemple devenu célèbre de ce principe. Il imagine un incendie dans la maison de Fénélon (1695-1715), dans laquelle l'auteur des *Aventures de Télémaque* (1699) est prisonnier des flammes avec sa femme de chambre [Godwin, 1793]. Vous passez devant la maison en feu mais vous ne pouvez sauver qu'une seule personne. Le principe d'utilité exige que vous sauviez Fénélon. Imaginez maintenant que la femme de chambre soit votre mère. Cette information doit-elle influencer votre décision ? Non, répond, Godwin. L'action morale doit pouvoir être impartialement considérée comme telle. Le résultat de la délibération morale ne peut dépendre de l'identité ni des intérêts personnels du passant.

S'il est aisé à justifier rationnellement, le principe d'impartialité est, cependant, souvent intuitivement difficile à défendre et pose donc la question de la valeur morale des intuitions. Le message de Godwin est qu'on a tort de se fier à nos intuitions (ici sauver sa propre mère). C'est également celui de Peter Singer, qui défend la primauté du principe d'impartialité dans la délibération morale sur celui de la proximité biologique, psychologique ou spatiale [Singer, 1979]. Selon Singer, l'intuition morale selon laquelle nous aurions une obligation morale de porter de l'aide à un enfant qui se noie devant nous mais pas à celui qui meurt de faim au Bangladesh, à un enfant auquel on est sentimentalement ou biologiquement lié mais pas à celui qui nous est inconnu, est

injustifiable [Singer, 1979]. La souffrance ou la mort d'un enfant, quel qu'il soit, est dramatique et doit être évitée si elle peut l'être. L'argument est naturellement valable pour tous les êtres humains, mais également pour Singer, on le verra, pour les animaux.

### *Impartialité et égalité*

L'utilitarisme se présente généralement comme une théorie promouvant l'égalité. Historiquement, les auteurs utilitaristes ont souvent milité pour une plus grande égalité politique, sociale, raciale et sexuelle. Pour un radical tel que Bentham, l'utilité des aristocrates ne compte pas davantage que celle de leurs employés ou servants. Bentham est également l'un des rares, y compris parmi les utilitaristes, à s'être attaqué au système colonial [Cot, 1996]. Devant le calcul d'utilité, chacun compte pour un et seulement pour un. Cela explique, également, que, selon Bentham, tous les plaisirs (à utilité égale) se valent, qu'ils soient intellectuels ou artistiques ou, au contraire, populaires et matérialistes. Dans certains de ses textes, dont beaucoup sont longtemps restés non publiés, il prône également l'égalité des hommes et des femmes, idée que John Stuart Mill, puis Henry Sidgwick, défendront ouvertement. Tous ces éléments font de Bentham l'un des philosophes les plus radicaux de sa génération. Mais il a été plus loin encore, plus loin qu'aucun de ses contemporains n'avait osé s'aventurer jusqu'alors. Dans un essai écrit en 1785, intitulé *Offences Against One's Self* et resté inédit jusqu'au vingtième siècle, il plaide ainsi pour la décriminalisation de l'homosexualité alors considérée comme un crime punissable de la peine de mort, peine à laquelle il est d'ailleurs opposé [Calvert, 2006 ; Schofield, 2014]. Mais c'est surtout en tant que défenseur des droits des animaux que Bentham est aujourd'hui célébré. Il annonce en cela les travaux de Peter Singer sur la libération des animaux [Kniess, 2019].

Tous les utilitaristes n'ont pas été aussi loin que Bentham. John Stuart Mill a ainsi eu des positions controversées relativement aux colonies, au vote pluriel ou à l'exemplarité des classes supérieures. Mais le fait qu'ils aient été aussi égalitaristes que leurs principes les autorisaient à l'être est une question secondaire. La question qui nous concerne ici est celle de savoir si la théorie utilitariste est égalitariste, et si elle l'est *incidemment* (comme effet

corollaire de la maximisation d'utilité) ou *nécessairement* (comme conséquence logique du principe d'impartialité).

Le principe d'égalité peut, tout d'abord, se justifier par l'exigence de maximisation d'utilité établie par la théorie utilitariste elle-même. Il est, en effet, possible de le préférer à tout autre système de répartition par simple souci d'efficacité. Dès lors qu'on admet le principe de décroissance de l'utilité marginale (i.e. l'utilité de la dernière unité consommée), la maximisation de l'utilité totale préconise logiquement la répartition égalitaire des ressources et/ou des richesses. Cela se comprend aisément. Le premier carré de chocolat donne toujours plus de plaisir que le second, le second plus que le troisième etc. jusqu'à ce qu'il ne procure plus aucun plaisir, voire nous dégoûte. De la même façon, l'utilité procurée par un euro supplémentaire est beaucoup plus élevée pour une personne démunie que pour un millionnaire. L'utilitarisme permet donc de justifier, au moins en théorie, les politiques sociales fortement redistributives.

Un autre élément psychologique, noté par Hare [Hare, 1989, p.195], contribue à faire de l'égalité un facteur potentiellement important de la maximisation d'utilité totale. L'inégalité de fortunes ou de revenus est généralement génératrice de ressentiments ou d'envie, y compris chez les individus nantis. Or, l'impact négatif de ces sentiments sur l'utilité totale des individus ne doit pas être sous-estimé. Une répartition égalitaire, ou du moins pas trop inégalitaire, des richesses est donc, dans l'état présent de nos sociétés, propice à la maximisation de l'utilité. Il serait, cependant, excessif d'aller jusqu'à affirmer que la maximisation de l'utilité implique une politique de redistribution strictement égalitaire. L'argument de la décroissance de l'utilité marginale, présenté ci-dessus, n'est vrai que *toutes choses égales par ailleurs*. En pratique, cependant, toute politique redistributive a des effets négatifs sur l'utilité de ceux qui sont imposés, effets qui peuvent dans certains cas être supérieurs au bénéfice généré par ailleurs. Ici encore doivent être pris en considération l'impact psychologique de ces prélèvements, notamment la démotivation, la frustration, le sentiment d'injustice ou d'oppression. Il est plus exact d'affirmer que l'utilitarisme justifie une application *modérée* du principe de redistribution.

L'adhésion au principe d'égalité dictée par l'exigence de maximisation d'utilité est, dans ce cas, fortement contingente et dépend directement de l'état d'esprit des individus. Elle ne



permet pas de mettre en place une politique égalitaire partout et à n'importe quelle condition, même s'il est théoriquement possible de justifier une perte temporaire d'utilité par un gain d'utilité futur plus important. Pour que l'utilitarisme puisse être interprété comme une doctrine égalitariste, celui-ci ne peut entièrement reposer sur le principe de maximisation du bien-être.

Le principe d'impartialité permet de garantir l'égal respect des intérêts individuels que le mécanisme de maximisation d'utilité, seul, ne garantit pas. L'égale considération des préférences des individus justifie alors (plutôt qu'est justifié par) le calcul d'utilité. Cette approche est défendue, notamment, par Harsanyi, Hare et Singer. Pour Harsanyi, l'égale considération des préférences individuelles est garantie par la dimension contractualiste ainsi que l'hypothèse d'équiprobabilité de son modèle [Harsanyi, 1984]. Chez Hare, elle est le produit du processus d'universalisation exigé par le niveau critique du raisonnement moral (celui des 'archanges') [Hare, 1981]. Enfin, chez Singer, l'égale considération des intérêts est posée comme un 'principe éthique de base' [Singer, 1979, p.18], à partir duquel il construit un modèle qui repose sur l'empathie [Holton et Langdon, 1999]. Dans tous les cas, ce principe n'implique pas un égal traitement des individus.

La distinction de ces deux types de justification du principe d'égalité, l'incidence et la nécessité, est en pratique souvent difficile, sinon impossible, à établir. Selon Kymlicka, cet amalgame, qu'il soit volontairement entretenu ou non, explique en partie, l'attrait important de cette théorie [Kymlicka, 2003, p.46]. Il explique aussi, en partie, sa fragilité.

### **Limites du principe d'impartialité**

#### *De l'impartialité à l'impersonnalisme*

Si l'impartialité est généralement présentée comme une condition de la morale, elle peut être critiquée pour son coût jugé parfois exorbitant pour l'individu. En 1973, le philosophe britannique Bernard Williams publie 'Une critique de l'utilitarisme' dans laquelle il accuse l'utilitarisme de ne pas respecter l'intégrité des personnes [Williams, 1973]. Une action n'a de valeur que par son impact net, quelle que soit la personne qui l'entreprend, les raisons pour les lesquelles elle est entreprise et les moyens qu'elle utilise. L'impartialité qu'exige

l'utilitarisme de la part des individus les empêche d'entretenir des relations humaines étroites – et nécessairement privilégiées – avec leurs amis ou leur famille. Elle rend *de facto* impossible toute vie 'ordinaire' et promeut un système dans lequel toutes nos actions sont dictées par un calcul de maximisation. L'erreur des utilitaristes, affirme également Peter Railton [Railton 1984], est d'avoir ignoré la dimension profondément 'aliénante' de leur philosophie morale pour les individus, au point de rendre leur comportement complètement impersonnel et comme détaché de tout attachement sentimental et social.

La critique de Williams et Railton fait écho à un autre argument avancé par le philosophe américain John Rawls dans sa *Theory of Justice* [Rawls 1971, traduit en 1987]. Rawls y accuse l'utilitarisme d'avoir confondu impartialité et impersonnalisme. La procédure morale qui garantit que les intérêts de chaque personne soient également considérés n'implique pas, pour les contractualistes, que le point de vue de la personne soit totalement éliminé. En d'autres termes, l'objectivité morale ne se réduit pas à une agrégation d'unités d'utilité isolées de la personne. Rawls récuse le raisonnement qui conduit à penser la société à l'image de l'individu et, ainsi, à justifier le sacrifice interpersonnel d'utilités. Au niveau individuel, il est possible – et dans certains cas souhaitable – de sacrifier l'utilité d'une période pour le bénéfice d'une autre ou de plusieurs autres. C'est ce prescrit la règle de la prudence, qui repose elle-même sur le principe de neutralité par rapport au temps [Brink 2011]. En redistribuant l'utilité entre les différentes périodes de sa vie, l'agent maximise son utilité totale. L'erreur de l'utilitarisme, nous dit Rawls, est de croire (et surtout de nous faire croire) que le même raisonnement puisse également s'appliquer entre les individus, et qu'une redistribution des utilités entre les agents est objectivement souhaitable. Ce faisant, il nie la spécificité et la séparabilité de la personne [Rawls, 1971, traduit en 1987, p.53], et fait de l'utilitarisme une théorie, non pas impartiale, mais impersonnelle.

L'impartialité désigne, on l'a vu, l'impossibilité de justifier un choix moral par les intérêts d'une ou de plusieurs personnes. L'impersonnalisme désigne, quant à lui, l'impossibilité de justifier un choix moral du point de vue d'une ou de plusieurs personnes. Une approche impersonnelle de l'éthique est donc une approche dans laquelle le concept de personne est superfétatoire. Les implications morales de ce transfert sont énormes pour les individus. Au-delà du sentiment d'aliénation qu'il peut provoquer, il peut également

légitimer un certain nombre de politiques de redistributions contestables ou d'atteintes à l'intégrité des personnes.

### *Prendre les droits des personnes au sérieux*

La théorie utilitariste est généralement considérée comme antinomique avec la reconnaissance des droits de l'homme. Cette idée reçue est partiellement vraie et elle a pour origine les attaques féroces de Bentham contre les droits naturels. Le fondateur de l'utilitarisme a ainsi qualifié les droits naturels de 'non-sens sur des échasses' (*non-sense upon stilts*). Il faut, cependant, être relativement prudent dans les conclusions à tirer de ces expressions, et il est important 1° de les replacer dans le contexte intellectuel de l'époque et 2° de ne pas réduire la pensée utilitariste à celle de Bentham. La critique de la théorie des droits naturel est certes un élément fondamental de sa philosophie, qui ainsi qu'on l'a vu, s'est partiellement développé en réaction au système jusnaturaliste et au système jurisprudentiel défendu par Blackstone. Bentham critique notamment la rhétorique et la méthode fallacieuse de la référence à la nature. Il dénonce également l'incohérence de la *Déclaration des droits de l'homme* de 1789, notamment l'idée selon laquelle les droits seraient à la fois le produit de la souveraineté populaire et opposables à la loi. Les droits sont, chez Bentham, positifs dans le sens où ils sont créés par l'institution politique, de manière à répondre au mieux aux circonstances du moment, et ils ont vocation à guider les hommes dans leur recherche du plus grand bien pour le plus grand nombre. Pour reprendre une terminologie moderne, on dira, avec Lyons, que Bentham reconnaît les droits légaux (*legal rights*) dans la mesure où ils sont conçus comme leviers de la maximisation d'utilité, mais qu'il ne leur reconnaît aucune force morale autre que celle qui leur est donnée par l'autorité de la loi, autrement dit qu'ils ne sont pas équivalents à des droits moraux (*moral rights*) [Lyons, 1984]. Notons que cette approche du droit ne fut pas partagée par tous les utilitaristes, et que John Stuart Mill, notamment, reconnut l'importance des droits moraux. La philosophie utilitariste n'en est pas moins restée, pour ses opposants, fondamentalement incompatible avec le respect absolu des droits individuels. Parmi les critiques contemporaines les plus influentes furent celles de Ronald Dworkin et de Robert Nozick.

Ronald Dworkin (1931-2013) est avec John Rawls (1921-2002) l'une des grandes figures du libéralisme américain, et l'un des plus importants penseurs de la théorie libérale du droit, qu'il construit à la fois en opposition d'une part à la philosophie du droit alors dominante, sorte de combinaison entre le positivisme juridique et l'utilitarisme, défendue par le H.L. Hart, et au constructivisme rawlsien d'autre part. Il publie *Taking Rights Seriously* en 1977 (traduit en français en 1995), ouvrage dans lequel il défend la primauté des droits sur l'utilité. Selon lui, la nature même des droits est de ne pouvoir être pondérés ou contrebalancés par d'autres considérations, y compris le bien général. Il s'oppose donc explicitement au raisonnement utilitariste, selon lequel les droits ne se justifient que par leur utilité publique, mais également peuvent être individuellement violés lorsque la situation le nécessite. Cette approche s'oppose notamment à la possibilité d'un utilitarisme des droits dans lequel les droits de certaines personnes pourraient être violés afin de minimiser l'ensemble des violations des droits des autres personnes. Une telle approche est, selon Dworkin, contestable y compris du point de vue de la philosophie utilitariste pour laquelle l'égalitarisme représente un atout majeur.

L'utilitarisme des droits, version 'corrompue' de l'utilitarisme [Dworkin, 1984, p.154], viole le droit à l'indépendance morale des individus en donnant moins de poids à certains individus ou à certaines catégories de personnes qu'à d'autres. Dworkin considère ainsi que les lois condamnant les pratiques homosexuelles, lois justifiées par la prise en compte des préférences externes des hétérosexuels, constituent une violation du principe selon lequel la société doit traiter tous les concitoyens de la même façon. Une société dont la politique est guidée par les principes utilitaristes doit donc être particulièrement vigilante à ce que les droits à l'indépendance morale de chacun soient respectés. C'est la raison pour laquelle, les droits des individus doivent toujours avoir priorité (*trump*) sur le bien de la communauté (tel qu'il est déterminé par l'utilitarisme standard). Lorsqu'elle pénalise fortement les intérêts des minorités, la maximisation du bien-être général contrevient au respect du principe d'égalité qui est au fondement même de nos sociétés. Bien que très influente, la critique de Dworkin peut apparaître assez conciliante avec le système utilitariste qu'elle ne rejette pas totalement par ailleurs. De ce point de vue, la critique de Nozick est beaucoup plus radicale.

Publié en 1974, *Anarchy, State and Utopia* de Robert Nozick (1938-2002) constitue aujourd'hui la pierre angulaire de la pensée libertarienne contemporaine [Nozick, 1974]. Comme tous les libertariens, Nozick est favorable à un Etat minimal et rejette toute idée de redistribution *entre* les individus. Il s'oppose, en effet, à toute interférence publique qui viendrait violer le sacro-saint principe de propriété de soi hérité de Locke. Pour Nozick l'homme est *entièrement* propriétaire de son corps et, par extension, des fruits de son travail ou de son talent. Nul, et encore moins l'Etat, n'a le droit de 'spolier' les hommes d'une partie de leurs revenus pour les redistribuer aux plus nécessiteux. Ce principe est absolu et ne peut admettre aucune exception. La philosophie de Nozick, qui s'inspire de la théorie du droit naturel lockienne, s'oppose donc frontalement au principe de maximisation de l'utilité qui implique que certains se sacrifient pour le bien de la communauté. Nul n'a l'obligation de se soumettre à une politique qui le forcerait à sacrifier son intérêt personnel au nom du bien d'une communauté, qui n'a par ailleurs aucune existence en-dehors de celle de ses membres.

Nozick développe une approche déontologique du droit fondée sur la notion de contrainte. Les droits ne sont pas, selon lui, le produit de conventions politiques. Ils précèdent et ont, par conséquent, la priorité sur toute constitution politique. A la différence de certains libéraux qui considèrent la non-violation des droits humains comme le but de l'Etat, Nozick préfère penser les droits comme des 'contraintes secondaires' (*side constraints*), c'est-à-dire comme des contraintes morales interdisant absolument la violation des droits d'une personne tierce dans le but de satisfaire un intérêt personnel. Elle s'inspire directement du principe kantien selon lequel on ne doit jamais traiter les individus simplement comme des moyens, mais toujours en même temps comme des fins. Le droit de propriété que chacun a sur sa propre personne, et par extension sur les fruits de son travail, ne peut être enfreint par aucune institution politique, y compris démocratique, car les droits des individus sont naturels, autrement dit, antérieurs à toute constitution politique ou contrat social.

*L'utilitarisme est-il vraiment égalitaire ?*

La thèse de l'égalitarisme de l'utilitarisme a été fortement contestée dans les années 70 et 80. Deux fortes objections théoriques ont notamment été développées. La première prend

la forme d'une expérience de pensée, connue sous le nom de monstre d'utilité. On la doit au philosophe libertarien américain Robert Nozick. Celui-ci imagine un personnage, qui loin de se rassasier de ce qu'il consomme, éprouve un plaisir toujours plus important au fur et à mesure qu'il consomme [Nozick, 1974, p.41]. Dans la théorie utilitariste, la répartition égalitaire des ressources se justifie notamment par la décroissance de l'utilité marginale des biens (ou de la monnaie). Otez cette hypothèse, comme c'est ici le cas avec le monstre d'utilité, et la justification par incidence de la répartition égalitaire des ressources s'effondre d'elle-même. Si l'on peut espérer que l'utilitarisme cautionne l'égalitarisme dans certains cas, il est aussi tout à fait possible, au moins théoriquement, qu'il justifie également le sacrifice de tous pour le confort de très peu, voire d'un seul.

La conclusion de Nozick est que la philosophie utilitariste n'est pas nécessairement conforme aux principes de l'égalité. L'argument de Nozick, bien qu'intellectuellement stimulant, pose néanmoins problème car il remet également en question toute justice distributive fondée sur la notion de besoin des individus.

La seconde objection au principe de l'égalitarisme utilitariste fait appel à la notion de préférences externes. On la doit au philosophe du droit Ronald Dworkin. L'utilitarisme se targue de prêter une égale considération à tous les êtres humains en donnant le même poids aux préférences de chacun. Mais cela suffit-il à ce que les individus soient, en pratique, traités sur un strict pied d'égalité ? Pas nécessairement. Il est important de distinguer ce que Dworkin appelle les 'préférences externes' des 'préférences personnelles' [Dworkin, 1977, p.234]. En l'absence d'une telle distinction, il existe un risque réel de surpondérer les préférences de certaines personnes au détriment de celles d'autres. Les préférences individuelles peuvent, en effet, soit avoir trait à des objectifs personnels à atteindre (dîner dans un restaurant chic, être accepté dans une école prestigieuse etc.) soit avoir trait aux objectifs des autres, les deux pouvant être étroitement liés entre eux. Ainsi, dans une Amérique ségrégationniste, il est possible qu'un étudiant souhaite intégrer une université prestigieuse (préférence personnelle) à condition, cependant, que celle-ci soit socialement et ethniquement très exclusive (préférence externe). Satisfaire ce type de préférence ne peut se faire qu'au détriment des préférences des pauvres et des minorités ethniques et contrevient donc au principe d'égalité. *A contrario*, la satisfaction de mes préférences individuelles peut être fonction de la satisfaction des préférences de

personnes pour lesquelles j'éprouve de la sympathie. Si j'ai envie que Marie obtienne une place dans l'université de son choix, la préférence (personnelle) de Marie « intégrer l'université X » sera alors comptée deux fois. Plus Marie sera aimée, plus sa préférence sera surpondérée par rapport à un étudiant lambda, et *a fortiori* par rapport à ceux qui font l'objet de préférences externes négatives.

Le principe selon lequel chacun ne compte que pour un et seulement pour un nous conduit donc à éliminer les préférences externes de l'agrégation des préférences. C'est d'ailleurs (en partie) la position d'Harsanyi qui s'oppose à l'intégration de toute préférence anti-sociale (telle que l'envie ou la jalousie) dans le calcul utilitariste [Harsanyi 1982]. Comment justifier, cependant, autrement que par un argument *ad hoc* que seules les préférences externes à effets négatifs soient exclues du calcul utilitariste, et qu'*a contrario* les préférences externes à valeur positive, telles que l'adhésion aux principes égalitaristes, la bienveillance ou l'héroïsme – qui sous-tendent la philosophie utilitaristes – soient, elles, systématiquement conservées ? Comment traiter, enfin, les préférences individuelles qui, bien que non externes, n'en sont pas moins critiquables d'un point de vue moral ? Que faire, par exemple, des préférences ultra-égoïstes des individus qui, si elles sont satisfaites, entraîneront nécessairement un accroissement des inégalités ? Pour Richard Hare [Hare, 1982], il n'y a pas de raisons *a priori* d'écarter ce type de préférences. Dans le cas où un arbitrage serait nécessaire pour départager les gagnants des perdants, on aurait recours à une expérience de pensée similaire à celle du voile d'ignorance, dans laquelle l'intérêt de chacun serait successivement évalué de son propre point de vue, et l'action maximisant l'utilité de tous les protagonistes, traités chacun sur un pied d'égalité, serait préférée (même si elle a pour conséquence de désavantager ceux qui sont déjà les plus démunis).

#### *Encadré n°4. Le système des enchères de Dworkin*

*Pour Dworkin, comme pour Sen, une répartition égalitaire des biens premiers (socialement construits) ne permet pas de rendre justice au concept d'égalité. Un égal respect entre les individus impose aussi de prendre en compte la répartition initiale de leurs dotations naturelles, que ce soit leurs capacités physiques ou intellectuelles, leurs talents ou leur inégale propension à être malade ou à mourir prématurément. L'inégalité de ces 'ressources personnelles' n'étant pas le fait des individus eux-mêmes, elle peut faire l'objet d'une compensation. En revanche, les inégalités produites par les différences de 'personnalité', c'est-à-dire par le caractère des individus, leurs préférences, leur ambition ou leur aptitude à l'effort, elles, ne peuvent faire l'objet d'une compensation.*

*Pour évaluer l'importance des compensations auxquelles certains individus ont droit, Dworkin imagine un système d'enchères associé à un calcul assurantiel assez complexe. Le système d'enchères permet de déterminer le panier de biens optimal pour chaque individu, c'est-à-dire selon sa théorie, celui qui ne sera préféré à aucun autre par l'individu l'ayant acheté. Le système assurantiel permet, quant à lui, de prendre en compte les handicaps naturels subis par certains individus, handicaps souvent coûteux et ne leur permettant pas de s'offrir le panier désiré. Pour illustrer son propos, Dworkin imagine une fiction dans laquelle des naufragés se retrouvent isolés sur une île déserte, et proposent de se répartir également les ressources à leur disposition. Plutôt que de diviser les biens en parts égales, ils préfèrent les répartir en tenant compte des préférences de chacun. Pour qu'une répartition différenciée mais néanmoins égalitaire soit réussie, il faut, selon Dworkin, qu'elle passe 'le test de l'envie', c'est-à-dire qu'au final chacun préfère le panier de biens qu'il a acquis à celui des autres. Une vente aux enchères des différents biens, dans laquelle chacun est doté d'un même pouvoir d'achat (dans l'exemple des naufragés, des coquillages), apparaît le plus approprié. On pourrait, cependant, objecter à ce système, que même dans la meilleure configuration possible, ce système n'est pas véritablement juste puisque, bien que toutes les personnes aient des ressources initiales égales (ici des coquillages), certaines d'entre elles ont des besoins spécifiques (liés à des handicaps ou à une mauvaise santé par exemple), besoins coûteux qui ne leur permettent pas d'obtenir le panier de biens qu'ils auraient pu s'offrir autrement.*

*Dworkin suggère alors d'intégrer au système d'enchère un mécanisme assurantiel, mécanisme par lequel les agents choisiraient (collectivement) un niveau de couverture contre le risque d'être naturellement défavorisé. Les individus seraient ainsi invités à se prononcer, derrière ce qui peut faire penser à un voile d'ignorance, sur la part de ressources qu'ils seraient prêts à sacrifier pour se couvrir contre les risques d'handicaps naturels. Une fois le taux de couverture choisi, tous paieraient la même prime, et les plus défavorisés dans la distribution des capacités naturelles seraient compensés à hauteur de leur handicap. A partir de ce schéma, il est possible de concevoir un système de redistribution financé par la part des ressources individuelles dévolue à compenser les inégalités naturelles (les impôts).*

### **(Ré)Intégrer la notion de personne**

Face aux difficultés soulevées par le principe d'impartialité, un certain nombre d'auteurs ont tenté de réintégrer la notion de personne dans le raisonnement conséquentialiste (voire utilitariste) : (1) en associant la notion de droit à celle d'utilité, (2) en prenant en compte la situation relative de la personne dans les politiques de redistribution, ou (3) en assouplissant le principe d'impartialité et en introduisant le concept de prérogative personnelle.

### *(Ré)Concilier utilité et droits*

Dans son article « Rights and Agency » [Sen, 1982], Amartya Sen présente une théorie tentant de réconcilier la notion de droits et celle d'utilité. Il considère que ni l'approche déontologique stricte, fondée sur la notion de contrainte absolue, ni l'approche



conséquentialiste, fondée sur le welfarisme, ne permet de répondre correctement à la complexité des situations réelles. Bien qu'antinomiques, elles partagent le même présupposé – qu'il juge erroné – à savoir le fait que le respect ou la violation des droits ne joue aucun rôle dans l'évaluation des états du monde. Or, selon lui, les droits des individus, ou plus exactement le fait qu'ils soient respectés, constituent également une finalité de la politique publique, et le fait qu'ils soient plus ou moins respecté doit faire partie de l'évaluation globale des différents états du monde pour déterminer quelle est la meilleure action à mener. Le respect des droits constitue un but, au même titre que peut l'être le bien-être individuel. Il appelle cela le « système des droits-buts » (*goal rights system*). Il définit, par ailleurs, les droits des individus comme des droits à exercer leur capacités (*capabilities*), plutôt que des droits à faire ou à ne pas être sujet à telle ou telle action.

A l'instar de Sen, Phillip Pettit défend l'idée qu'un conséquentialisme restrictif reposant sur la reconnaissance des droits de la personne (*rights-restrictive consequentialism*) est supérieure aux formes régulières du conséquentialisme [Pettit, 1988]. Pour Pettit, qui a par ailleurs développé une théorie de la liberté comme non-domination [Pettit, 1997], les droits ont nature à protéger les individus plutôt qu'à promouvoir leurs intérêts. Ils sont inaliénables et ne peuvent faire l'objet d'échanges ou de négociation. Enfin, ils doivent être spécifiques. L'utilité des droits ne se limite donc pas au fait que nos intérêts (notre intégrité physique, nos propriétés etc.) ne soient pas attaqués mais au réconfort que l'on tire quotidiennement du fait qu'on les sait être protégés. Or, cet état, et l'utilité qu'il génère, n'est possible que dans le cadre d'un conséquentialisme restrictif (cf chapitre 3), dans lequel l'agent accepte de contraindre le processus de maximisation lorsque les droits d'un individu sont menacés. Dès lors que l'agent maximisateur a la possibilité d'ignorer le droit des individus (et que cela se sait), même s'il n'en use pas, l'utilité générée par la nature protectrice des droits disparaît. La maximisation d'utilité implique donc que le processus de maximisation des agents soit contraint par la reconnaissance de leurs droits.

### *Le prioritarisme*

L'utilitarisme est-il injuste envers les plus démunis ? Le raisonnement conséquentialiste peut-il tourner à l'avantage des plus favorisés ? Le plaisir procuré par l'accumulation de

richesses ne semble, en effet, pas avoir de limites, ce qui confirmerait l'hypothèse selon laquelle la satisfaction d'être riche échapperait à la loi de la décroissance de l'utilité marginale. En soi, cette exception pourrait à elle seule donner raison à Nozick et remettre en cause l'égalitarisme déclaré de l'utilitarisme. Mais la faille de la philosophie utilitariste est peut-être plus importante encore lorsque celle-ci repose sur la théorie des préférences plutôt que sur l'hédonisme (cf chapitre 2).

La formation de nos préférences n'est, en effet, pas totalement indépendante du contexte dans lequel celles-ci sont formées. Il semble ainsi assez rationnel pour un individu de former ses préférences en fonction de la probabilité qu'il a ou qu'il estime possible de les satisfaire. Galbraith considère même qu'il s'agit là d'une réponse naturelle à la pauvreté » [Galbraith, 1979]. De plus, et ainsi que le philosophe Jon Elster l'a montré, il est également possible que nous adaptions progressivement nos préférences à notre état de manière à éviter le sentiment de frustration qu'une préférence non satisfaite ne manquerait pas de provoquer. Elster prend l'exemple du renard de la fable de La Fontaine ('Le Renard et les Raisins', elle-même inspirée d'Esopé), qui, affamé, repère une grappe de raisins mûrs, essaie en vain de l'atteindre et décide qu'ils sont verts et 'bons pour les goujats' [Elster, 1983, p.109]. Si tel est le cas, la satisfaction des préférences individuelles devient totalement déconnectée de leurs besoins réels. Pire, une répartition des ressources peut, sur la base des préférences individuelles, privilégier les mieux dotés aux plus démunis, et ainsi contribuer à accroître encore l'inégalité des richesses.

La théorie de la justice de Rawls, qui repose sur une théorie objective de la valeur mais également sur le principe de différence (selon lequel les inégalités économiques et sociales doivent être telles qu'elles soient au plus grand bénéfice des plus désavantagés), est censée répondre à cette objection. Elle implique, cependant, le rejet de l'utilitarisme. Le prioritarisme se présente comme une approche hybride entre l'exigence de maximisation de l'utilité totale, dont les implications peuvent être très défavorables aux plus démunis, et le respect absolu de l'intérêt particulier qui interdit tout principe de redistribution. Cette théorie se différencie néanmoins du principe de différence de Rawls qui interdit toute redistribution qui aurait un impact minime (mais néanmoins négatif) sur les plus défavorisés y compris si celui-ci se traduit par un immense bénéfice pour le reste de la communauté. Bien que l'utilité des plus démunis doive, selon les partisans du prioritarisme,

faire l'objet d'une attention toute particulière de la part des décideurs publics, celui-ci doit néanmoins pouvoir être mis en balance avec l'utilité totale. Il est ainsi possible de maximiser l'utilité totale tout en donnant la priorité aux intérêts des plus démunis.

Pour Derek Parfit, cette approche peut se justifier par 'la décroissance de l'importance morale marginale de l'utilité des personnes', autrement dit par le fait qu'une égale amélioration de l'utilité de plusieurs personnes n'a pas la même importance morale selon que ces personnes aient une utilité plus ou moins grande. Selon Parfit, des avantages identiques (en termes d'utilité) ne devraient pas systématiquement avoir le même poids [Parfit, 1997, p.213]. Cette théorie présente néanmoins un certain nombre de difficultés techniques liées à la mesure des améliorations respectives des situations personnelles en l'absence de toute comparaison interpersonnelle d'utilité [Otsuka, 2015]. Elle ravive également la critique faite depuis Rawls à l'utilitarisme classique, critique selon laquelle l'utilitarisme ne prend pas suffisamment la séparabilité des personnes au sérieux [Otsuka et Voorhoeve, 2009].

### *Existe-t-il des prérogatives personnelles en morale ?*

Les objections faites au principe d'impartialité sont fortes et fragilisent considérablement le discours utilitariste classique. Dans les chapitres précédents, nous avons évoqué plusieurs types de réponses apportées par les utilitaristes pour en amoindrir les effets, notamment l'excessive exigence morale. Une première forme de réponse consiste à abandonner l'objectif de la maximisation de l'utilité nette, en adoptant par exemple une version modérée de l'utilitarisme (*satisficing*). Cela permet ainsi de justifier les comportements des ménages qui donnent 'un peu' aux organismes caritatifs mais gardent l'essentiel de leur revenu pour l'entretien et l'éducation de leur famille. Une seconde forme de réponse a été apportée par l'utilitarisme de la règle, ou plus exactement par le caractère indirect de l'optimisation de l'utilité qu'il permet. La thèse sous-jacente est qu'il est plus efficace que chaque parent s'occupe de ses enfants et que les amis prennent soin d'eux entre eux, plutôt que nous soyons tous individuellement et impartialement responsables du bien-être des autres [Hare, 1981]. Cette théorie argue qu'elle ne viole pas vraiment le principe d'impartialité puisque cette forme de partialité est généralisable et, de fait,

généralisée. Elle reste, cependant, pour certains auteurs, problématique dans la mesure où les relations communautaires sont souvent génératrices de préjugés (religieux, nationaux, sociaux, sexistes, raciaux etc.), qui peuvent être ensuite utilisés pour exclure les minorités de l'aide dont ils auraient pourtant prioritairement besoin [Gruen, 1999].

Le troisième type de réponse est beaucoup plus radical puisqu'il consiste à (partiellement) abandonner le principe de l'impartialité et à lui préférer une approche personnelle (*agent-centered*) [Mulgan 2006 ; Scheffler 1982]. Dans *The Rejection of Consequentialism*, Samuel Scheffler, l'un des principaux partisans d'un conséquentialisme modéré, prend acte du peu d'importance que l'utilitarisme accorde aux personnes, et à leurs désirs [Scheffler, 1982]. Il défend une approche de la morale recentrée sur la personne (*morality of personal concerns*). Scheffler compare deux alternatives possibles au conséquentialisme. La première repose sur les conceptions totalement centrées sur l'agence (*fully agent-centered*) et désigne implicitement la théorie déontologique. La seconde est une conception hybride qui combine conséquentialisme et reconnaissance morale du statut de l'agent. En intégrant la prérogative de l'agent (*agent-centered prerogative*), la théorie hybride permet de prendre en compte l'indépendance du point de vue de la personne.

Dans les théories conséquentialistes pures, telles que l'utilitarisme, le point de vue de la personne n'est pas complètement ignoré. L'utilité ou la désutilité d'une action pour la personne est prise en compte dans le calcul global d'utilité, mais lorsque celle-ci ne représente qu'une part infime du total, elle est *de facto* négligée. La prérogative de l'agent octroyée par la théorie hybride, au contraire, donne un poids *moral* à l'agent, et ce quelle que soit la part de son utilité dans l'utilité totale. Elle reconnaît l'indépendance morale de celui, et l'importance de ses projets ou engagements personnels. Elle permet donc à l'agent d'échapper au calcul impersonnel du maximum d'utilité. Scheffler appelle cela la stratégie de libération (*the liberation strategy*). Elle est opposée à la stratégie de la maximisation (*the maximization strategy*), qui exige que chaque agent agisse toujours de manière à maximiser le nombre de personnes qui réussissent leurs projets de vie [Scheffler, 1982, p.59-60]. La stratégie de la maximisation reconnaît l'importance de l'indépendance du point de vue personnel dans la mesure où celle-ci contribue au bien individuel, mais elle ne lui donne pas une préséance morale. Scheffler entend ainsi démontrer, contre le

déontologisme, qu'il est possible de justifier le sacrifice de l'intérêt d'une personne au profit d'une autre, sans remettre en cause l'intégrité des personnes [Spitz, 2013].

Tim Mulgan s'inscrit dans cette perspective générale mais l'applique plus spécifiquement aux questions intergénérationnelles [Mulgan, 2006]. Face aux enjeux posés par la surpopulation dans le futur, il considère ainsi que seule une approche centrée sur la personne permet de justifier le fait de pouvoir se reproduire aujourd'hui. L'approche personnelle du conséquentialisme ne justifie pas, cependant, l'égoïsme. Elle s'accompagne également d'obligations qui nous contraignent à justifier nos actions aux personnes qui en seront victimes. Selon Mulgan, cela donne à nos obligations envers les générations futures une dimension d'urgence que le raisonnement impersonnel est incapable de transmettre [Mulgan, 2018, p.546].

### **L'impartialité en pratique : l'utilitarisme face à l'éthique appliquée**

Malgré les difficultés qu'elle soulève, la règle de l'impartialité reste un principe de base du raisonnement utilitariste. Son importance apparaît pleinement dans les grands problèmes éthiques auxquels nous faisons face. Sans explicitement le remettre en cause, nous avons parfois tendance à limiter l'étendue de ce principe, et ainsi à limiter nos obligations morales envers les autres. L'éthique appliquée remet en perspective les implications réelles d'un tel positionnement. Nous nous concentrerons sur trois cas, illustrant chacun une forme d'impartialité, les famines (égalité entre les hommes), la souffrance animale (égalité entre les hommes et les animaux) et le réchauffement climatique (égalité entre les générations du vivant).

#### *Les famines et la grande pauvreté*

En 1972, Peter Singer, suite à un terrible épisode de famines au Bangladesh, publia un article, devenu classique, dans lequel il défend une approche strictement impartiale de la morale (conséquentialiste) [Singer, 1972]. Selon Singer, il n'y a, en effet, aucune différence entre le fait de ne pas sauver un enfant qui se noie sous nos yeux dans une mare peu profonde et le fait de ne pas sauver un enfant qui meurt de faim à plusieurs milliers de

kilomètres. Dans les deux cas, nous avons la possibilité de sauver la vie d'un enfant à peu de frais. La distance géographique ou la proximité culturelle ne peuvent justifier un traitement différencié entre ces deux enfants. Dans les deux cas, nous avons une obligation morale de sauver ces enfants. Se soustraire à cette obligation, pour ne pas tremper ses chaussures ou perdre quelques dollars, constitue une grave faute morale.

L'article de Singer a été très commenté, et a joué un rôle important dans le développement de l'éthique appliquée et le renouveau de l'utilitarisme. Parmi les critiques les plus souvent avancées, citons celle de l'exigence morale (*moral demandingness*), celle de la pertinence de la notion de distance en morale [Kamm, 1999], ou celle qui consiste à refuser la notion de responsabilité négative en morale. Selon cette dernière objection, si les individus ont effectivement une obligation morale de ne pas sciemment nuire aux autres, il n'existe, en revanche, aucune obligation morale à soulager les souffrances existantes. Il est louable de vouloir soulager la misère du monde, mais ne pas le faire ne peut constituer une faute morale. Les individus ne sont moralement responsables que des conséquences des actions qu'ils réalisent, et non pas de celles qu'ils auraient pu mettre en œuvre. Cette objection conduit Colin McGinn à justifier le véganisme prôné par Singer (nous sommes responsables des souffrances des animaux que nous consommons ou exploitons), mais non pas ses appels à l'aide internationale pour lutter contre la grande pauvreté [McGinn, 1999]. Ce dernier argument est, cependant, contestable. Tout d'abord, il n'est pas certain que, collectivement, nous ne soyons pas indirectement responsables de la grande pauvreté au niveau mondial [Pogge, 2002]. Ensuite, à supposer que l'on n'ait aucune responsabilité morale dans le malheur d'autrui, comme c'est le cas pour l'enfant qui se noie dans la mare, cela ne nous affranchit pas de toute obligation morale de lui porter secours. L'analogie de Singer est donc recevable. Reste donc l'argument de la distance et celui de l'exigence morale, qui menacent le principe d'impartialité.

Nous préférons tous porter de l'aide à nos proches plutôt qu'à des étrangers. Nous préférons tous soulager les souffrances dont nous sommes personnellement témoins plutôt que celles que l'on nous a rapportées. Mais avons-nous raison ? Devons-nous toujours faire confiance à nos intuitions morales ? Dans les chapitres précédents, nous avons présenté certaines évolutions théoriques dont l'utilitarisme a fait l'objet au vingtième siècle. La plupart de ces évolutions avaient pour but de composer avec les

exigences morales de cette théorie. L'éthique appliquée, et notamment la souffrance des êtres humains, mais aussi des animaux, vient nous rappeler que la validité d'une théorie morale ne se juge pas par le niveau de confort qu'elle nous autorise. Une 'bonne' théorie morale doit avant tout être capable de répondre aux grands enjeux éthiques qui se posent à elle, et la détresse alimentaire d'une partie de la population mondiale en constitue certainement un.

### *L'éthique animale*

Le principe de l'impartialité nous enjoint à ne jamais prendre personnellement parti dans une délibération morale et à considérer toutes les parties affectées comme ayant un statut égal. Selon l'expression de Bentham, « chacun compte pour un et uniquement pour un ». Un certain nombre de penseurs, dont Bentham lui-même, se sont cependant posé la question de savoir si cette impartialité ne s'appliquait pas également à *toutes* les parties affectées, y compris les animaux [Kniess, 2019]. Dans une note de son *Introduction aux principes de morale et de législation*, Bentham remarque ainsi que le critère hédoniste retenu implique que les parties affectées par une action sont logiquement celles qui sont en capacité d'en souffrir. Or la souffrance n'est pas une spécificité humaine : « Pourquoi [leurs intérêts] ne sont-ils pas, universellement, tout autant que ceux des créatures humaines, considérés en fonction des différences de degré de sensibilité ? [...] Autrefois, et j'ai peine à dire qu'en de nombreux endroits cela ne fait pas encore partie du passé, la majeure partie des espèces, rangée sous la dénomination d'esclaves, étaient traitées par la loi exactement sur le même pied que, aujourd'hui encore, en Angleterre par exemple, les races inférieures d'animaux. Le jour viendra peut-être où il sera possible au reste de la création animale d'acquiescer ces droits qui n'auraient jamais pu lui être refusés sinon par la main de la tyrannie.[...] La question n'est pas « peuvent-ils raisonner? », ni « peuvent-ils parler ? », mais « peuvent-ils souffrir? ». » [Bentham, 1789, traduit en 2011, p.324-25].

Longtemps restée secondaire, cependant, l'intégration de la souffrance animale dans le calcul de maximisation a été sérieusement prise en compte à partir des années 1970, à nouveau sous l'impulsion de Peter Singer. Parallèlement à sa lutte contre la pauvreté et les famines dans le monde, Peter Singer s'est en effet fait connaître pour son engagement

précoce pour la cause animale. Son premier livre, *Animal Liberation*, initialement publié en 1975, a joué un rôle moteur dans la lutte contre les expérimentations, l'exploitation et la consommation des animaux. Ses travaux ont eu une influence importante et immédiate. Inspiré par Singer, l'activiste Henry Spira a ainsi réussi à mobiliser l'opinion publique américaine contre l'expérimentation sur les animaux dans l'industrie cosmétique, en menant, dans les années 1970 et 1980, des campagnes publicitaires très préjudiciables pour les marques [Singer, 1998, traduit en 2018].

L'argument de Singer repose sur l'analogie entre la discrimination sexuelle ou raciale et la discrimination spéciste. Le spécisme (il est un des premiers à employer et diffuser cette expression) est, pour lui, tout aussi condamnable que le sexisme ou le racisme car il viole le principe d'impartialité : « Le spécisme est un préjugé ou une attitude de parti pris en faveur des intérêts de sa propre espèce et à l'encontre des intérêts des membres des autres espèces » [Singer 1975, traduit en 2012, p.73]. En plus de l'abandon (ou de la stricte limitation) des expérimentations animales, Singer exhorte ses lecteurs à prendre conscience de leur responsabilité, en tant que consommateur, dans la maltraitance des animaux, et, par conséquent, à cesser de consommer des produits d'origine animale.

Ce type de défense de la cause animale (il est difficile de parler de droits des animaux dans une perspective utilitariste) fait généralement l'objet d'une double attaque : elle est, tout d'abord, remise en cause par ceux qui considèrent qu'elle va 'trop loin' (en exigeant des hommes qu'ils changent leurs habitudes alimentaires, par exemple), mais est aussi fortement critiquée par ceux qui considèrent qu'elle ne va pas 'assez loin', en refusant d'accorder aux animaux de véritables droits.

L'utilitarisme est, en effet, un instrument à double tranchant pour les militants de la cause animale. S'il permet de prendre en compte la souffrance des animaux, ou leur préférence à rester en vie, il peut également justifier ces mêmes souffrances (ou leur mise à mort) au nom de l'utilité totale. Singer a ainsi été attaqué pour ne pas systématiquement rejeter les expérimentations animales, notamment celles considérées comme nécessaires pour la mise au point de traitements médicaux pour l'homme [Gruen, 2016]. L'utilitarisme permet également de justifier la consommation et l'exploitation animale lorsque les conditions d'élevage des animaux sont jugées compatibles avec le bien-être animal. Selon l'argument



dit du remplacement, des millions de vies animales – et l'utilité qui leur est associée – n'existeraient simplement pas sans l'élevage. Lorsque ces animaux sont tués, ils sont remplacés par d'autres. Richard Hare a proposé de recourir à l'indicateur QALY (*Quality-Adjusted Life Year*), généralement utilisé en médecine pour évaluer l'utilité d'une intervention, pour juger de la légitimité du régime végétarien. Selon lui, l'objectif de la maximisation QALY est difficilement compatible avec l'abandon complet de la consommation animale [Hare, 1999, p. 240]. L'adoption d'un régime 'demi-végétarien' semble plus adapté. Il est également plus réaliste car moins exigeant.

Singer refuse 'empiriquement' l'argument du remplacement, même si ses écrits ne permettent pas vraiment de le remettre théoriquement en cause [Delon, 2016]. L'utilitarisme, on l'a vu, est multiforme et peut conduire à des résultats très différents selon le modèle théorique, la métrique ou le type d'agrégation adoptés. Prendre en compte le bien-être animal ne signifie pas, *a priori*, pondérer également les préférences humaines et les préférences animales [Kagan, 2019]. Une version 'modérée' de l'utilitarisme (*satisficing*) pourrait également se satisfaire d'une réduction des souffrances animales sans nécessairement remettre en cause le système d'élevage intensif. Enfin, selon la théorie du code moral idéal, pour être optimifuge, une règle morale doit pouvoir être internalisée relativement aisément. Plus la règle est contraignante, plus son enseignement est coûteux, moins efficace elle est.

Certains jugeront que la flexibilité de l'utilitarisme constitue l'une de ses principales limites, d'autres y verront un atout. La souffrance animale n'est, cependant, pas la seule considération à prendre en compte dans le calcul utilitariste. L'exploitation animale a de nombreuses autres conséquences sur l'utilité totale. Selon l'Organisation des Nations unies pour l'alimentation et l'agriculture (FAO), environ un quart des terres émergées est mis en pâturage et un tiers de la production agricole sert à l'alimentation animale<sup>243</sup>. Au total, plus de deux tiers des terres agricoles seraient directement ou indirectement utilisées pour la production animale. Dans le même temps, environ huit cents millions de personnes souffrent de malnutrition dans le monde et plus encore ont des difficultés à acheter des céréales, dont le cours est maintenu artificiellement élevé par la demande liée à

---

<sup>243</sup> <http://www.fao.org/3/ar591e/ar591e.pdf>

l'alimentation animale. Le recours à l'agriculture intensive et à la déforestation massive pour augmenter la production céréalière a, de son côté, d'importantes conséquences négatives sur l'environnement. Il est donc relativement aisé de démontrer, d'un point de vue utilitariste, qu'une réduction, voire l'abandon total, de la consommation animale profiterait au bien-être des animaux comme des humains. Toute la difficulté, cependant, est de savoir si cela suffit pour constituer une obligation orale.

### *L'éthique environnementale*

*Last but not least*, les théories utilitaristes nous fournissent les outils conceptuels pour nous interroger sur notre devoir moral face à la réalité et aux conséquences probables de nos actions sur le réchauffement climatique. Dans son rapport de 2019, le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) insiste sur la nécessité de respecter les objectifs d'émission de gaz à effets de serre approuvés lors de l'Accord de Paris de 2015. Les experts mettent notamment les gouvernements en garde contre la fonte des surfaces glacées, l'élévation du niveau des océans, et le bouleversement des écosystèmes terrestres et marins anticipés. Certaines zones géographiques pourront être submergées, d'autres pourraient se désertifier, d'autres enfin feront l'objet de catastrophes climatiques (tornades, cyclones, canicules etc.) de façon plus régulière et plus intense. Toutes les populations verront leurs conditions de vie impactées, mais certaines, en général celles des pays du Sud, seront plus durement affectées que les autres.

La dimension collective de la catastrophe annoncée semble a priori constituer un facteur favorable au principe d'impartialité. Certains éléments cités ci-dessus en compromettent l'exécution. Certaines régions du monde devraient, tout d'abord, être plus affectées que d'autres. Les générations futures seront, ensuite, plus durement frappées que les générations présentes. Enfin, toutes les espèces animales ne seront pas menacées de la même manière. Les humains de la présente génération et vivants dans les pays industrialisés du Nord seront, en moyenne, mieux protégés que les autres. Ce sont,

pourtant, les seuls en position de contrer le phénomène. Une remise en cause du principe d'impartialité peut, dès lors, avoir des effets irréversibles et dramatiques.

Les débats théoriques présentés dans cet ouvrage, s'ils peuvent parfois paraître spéculatifs, prennent ici toute leur importance. Au-delà de la question de l'impartialité, l'utilitarisme pose un certain nombre de questions auxquelles nous devons répondre si nous voulons nous engager dans une réflexion morale rigoureuse sur le futur de l'humanité : Devons-nous faire le *maximum* pour éviter le scénario avancé, au risque de sacrifier l'essentiel de notre confort de vie, ou pouvons-nous nous contenter d'un effort louable (*utilitarisme modéré*) ? Avons-nous réellement une obligation de maximiser le bien-être des générations futures ou pouvons-nous nous contenter de minimiser leurs souffrances à venir (*utilitarisme négatif*) ? Devons-nous calculer l'utilité des futurs agents en nous fondant sur nos préférences actuelles ou ce que pourraient être leurs préférences dans un lointain futur (*utilitarisme des préférences*) ? Devons-nous assumer une responsabilité morale individuelle face à cette sombre perspective ou pouvons-nous répartir les efforts à consentir entre tous (*utilitarisme indirect*) ? Devons-nous établir un code moral, facilement assimilable, qui s'applique à tous tout le temps, plutôt que de laisser les individus délibérer sur la légitimité de leurs actions (*utilitarisme de la règle*) ? Devons-nous répartir les sacrifices de façon inégale pour éviter que les plus pauvres paient un prix disproportionné pour une situation dont ils ne sont que très faiblement responsables ? etc. Il n'y a naturellement pas de réponse unique à ces questions. Les chapitres précédents comportent, cependant, un certain nombre d'arguments qui devrait permettre à chacun de se faire une opinion. Chaque question introduit un nouveau niveau de difficulté dans le raisonnement et nous rapproche un peu plus de la complexité du dilemme moral auquel nous devons faire face. Cette complexité ne doit pas nous décourager. L'enjeu est trop important.

## Conclusion

L'utilitarisme, on a pu le constater, est un ensemble théorique très divers et beaucoup plus complexe qu'on a parfois pu le présenter. Sa principale force est d'avoir pu s'adapter aux évolutions théoriques et sociales des deux derniers siècles. L'adoption des préférences individuelles comme mesure de l'utilité ainsi que de la règle comme objet de l'évaluation ont ainsi constitué ses deux principales avancées théoriques majeures au vingtième siècle. L'utilitarisme est également une philosophie, qui de par sa nature conséquentialiste, a une forte dimension pratique. Cela a certainement joué en sa faveur, et il fait peu de doute que le renouveau de l'utilitarisme à partir des années 70 doit beaucoup à sa capacité à nourrir les débats de l'éthique appliquée.

Aujourd'hui, les nouveaux enjeux de l'éthique nous invitent à faire des arbitrages entre la recherche de notre bien-être et la détérioration de l'environnement, la malnutrition d'une partie de la population mondiale, l'extinction de certaines espèces, la souffrance animale ou encore le bien-être des générations à venir. Il s'agit de choix difficiles mais nécessaires. La philosophie utilitariste, malgré ses limites, peut nous aider à trouver des réponses. La gravité des enjeux et la complexité des questions nous interdit, cependant, de céder à la facilité, à l'idéologie et aux stéréotypes. De leur côté, les philosophes s'emploient à améliorer leurs théories afin qu'elles répondent mieux à l'évolution des questions éthiques. Il nous appartient de nous informer de ces développements et de les exploiter au mieux pour pouvoir répondre aux fantastiques enjeux auxquels nous sont confrontés aujourd'hui.

## REPÈRES BIBLIOGRAPHIQUES

- ADAMS, R. M. [1976], "Motive Utilitarianism", *The Journal of Philosophy*, vol.73, n°14, p.467-481.
- ALBEE, E. [1902], *A History of English Utilitarianism*, Swan Sonnenschein & Co., Londres.
- ANDERSON M. et ANDERSON S. L. (eds), [2011], *Machine Ethics*, Cambridge University Press, Cambridge.
- ANSCOMBE, G.E.M. [1958], "Modern Moral Philosophy", *Philosophy*, Vol. 33, n° 124, p. 1-19.
- ARNESON, R. [1990], "Primary Goods Reconsidered", *Noûs*, vol.24, n°3, p.429-454.
- ARRHENIUS, G., RYBERG, J. et TÄNNSJÖ, T. [2017], "The Repugnant Conclusion", *The Stanford Encyclopedia of Philosophy*, ZALTA E.N. (ed.), URL = [<https://plato.stanford.edu/archives/spr2017/entries/repugnant-conclusion/>](https://plato.stanford.edu/archives/spr2017/entries/repugnant-conclusion/).
- ARROW, K.J. [1951], *Social Choice and Individual Values*, John Wiley & Sons, New York.
- AUDARD, C. [1999], *Anthologie historique et critique de l'utilitarisme*, 3 vols., Presses Universitaires de France, Paris.
- AUDARD, C. [2011], "Un utilitarisme libéral est-il un oxymore ? Hayek et la critique du 'constructivisme' », in Bozzo-Rey, M. et Dardenne, E. (eds.), *Deux siècles d'utilitarisme*, Presses universitaires de Rennes, Rennes, p. 171-192.
- AYER, A. J. [1972], "The Principle of Utility", in *Philosophical Essays*, Palgrave Macmillan (UK), London, p. 250-270.
- BACCELLI J., et MONGIN P. [2016], "Choice-Based Cardinal Utility", *Journal of Economic Methodology*, vol.23, n°3, p. 268-288
- BAUJARD, A. [2017], "L'économie du bien-être est morte. Vive l'économie du bien-être !" G. Campagnolo et J.-S. Gharbi (eds.), *Philosophie économique. Un état des lieux*, Editions Matériologiques, Paris, pp. 77-128.
- BELL, D. [2010], "John Stuart Mill on Colonies", *Political Theory*, vol.38, n°1, p.34-64.
- BENTHAM, J. [1840], *Oeuvres de Jérémie Bentham: essai sur la situation politique de l'Espagne, défense de l'usure. De l'organisation judiciaire et de la codification*, Société Belge de Librairie, Bruxelles.
- BENTHAM, J. [1789], *Introduction aux principes de morale et de législation*, traduction par le centre Bentham [2011], Vrin, Paris.
- BENTHAM, J. [1834], *Déontologie ou Science de la morale*, Charpentier, Paris.
- BENTHAM, J. [1968-2016], *The Collected Works of Jeremy Bentham*, ed. J. H. BURNS (1961-79), J. R. DINWIDDY (1978-83), F. ROSEN (1983-95), F. ROSEN and P. SCHOFIELD (1995-2003), P. SCHOFIELD (2003-), Athlone Press and Clarendon Press, London and Oxford.
- BERLIN, I. [1969], "John Stuart Mill and the Ends of Life" in *Four Essays on Liberty*, Oxford University Press, Oxford, p.173-206.

- BLAMIRE C., [2008], *The French Revolution and the Creation of Benthamism*, Palgrave Macmillan, Basingstoke.
- BONICCO, C. [2011], "L'utilitarisme idéal de Moore: un conséquentialisme non hédoniste." In M. Bozzo-Rey et E. Dardenne (eds.), *Deux siècles d'utilitarisme*, Presses Universitaires de Rennes, Rennes, p. 75-88.
- BOZZO-REY, M. [2011], "Langage et méta-éthique: l'utilitarisme de Hare en question", in BOZZO-REY, M. et DARDENNE E., *Deux siècles d'utilitarisme*, Presses Universitaires de Rennes, Rennes, p.89-99.
- BOZZO-REY, M. et DARDENNE E. [2011], *Deux siècles d'utilitarisme*, Presses Universitaires de Rennes, Rennes.
- BRADLEY, B. [2006], "Two Concepts of Intrinsic Value", *Ethical Theory and Moral Practice*, vol. 9, n°2, p.111-130.
- BRANDT, R.B. [1959], *Ethical Theory: The Problems of Normative and Critical Ethics*, Prentice-Hall, Englewood Cliffs, (N. J., USA).
- BRANDT, R.B. [1979], *A Theory of the Good and Right*, Clarendon Press, Oxford.
- BRANDT, R.B. [1992], *Morality, Utilitarianism, and Rights*, Cambridge University Press, Cambridge.
- BRINK, D.O. [1988], "Sidgwick's Dualism of Practical Reason", *Australasian Journal of Philosophy*, vol. 66, n°3, p.291-307.
- BRINK, D. [2010], "Mill's Ambivalence about Rights", *Boston University Law Review*, vol.90, p.1669-1704.
- BRINK, D.O. [2011], "Prospects for Temporal Neutrality" In Callender C. (ed.), *The Oxford Handbook of Philosophy of Time*, Oxford University Press, Oxford and New York, p.353-381.
- BRINK, D.O. [2013], *Mill's Progressive Principles*, Clarendon Press, Oxford.
- BRINK D. [2018], "Mill's Moral and Political Philosophy", in Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/win2018/entries/mill-moral-political/>.
- BROAD, C.D. [1916], "On the Function of False Hypotheses in Ethics," *International Journal of Ethics*, vol.26, n°3, p.377-97.
- BROAD, C.D. [1930], *Five Types of Ethical Theory*, Harcourt, New York.
- BROCK, D. [1993], *Life and Death: Philosophical Essays in Biomedical Ethics*, Cambridge University Press, Cambridge.
- BROOME, J. [1993], "A cause of preference is not an object of preference", *Social Choice and Welfare*, vol.10, n°1, p.57-68.
- BROOME, J. [2006], *Weighing Lives*, Oxford University Press, Oxford.
- CAILLE A., LAZZERI, C. et SENELLART, M. [2001], *Histoire raisonnée de la philosophie morale et politique. Le bonheur et l'utile*, Editions La Découverte, Paris.
- CALVERT, B. [2006], "Bentham and the Death Penalty", *Dialogue. Canadian Philosophical Review / Revue canadienne de philosophie*, vol.45, n°2, p.211-23
- CANTO-SPERBER, M. [1994], *La philosophie morale britannique*, Presses universitaires de France, Paris.
- CAPALDI, N. [2004], *John Stuart Mill: A Biography*, Cambridge University Press, Cambridge.
- CLAYES, G. [2013], *Mill and Paternalism*, Cambridge University Press, Cambridge.
- CLEMENTZ, F. [2004], "Prescription, motivation et rationalité", in Goffi, J.-Y. (ed.), *Hare et la philosophie morale*, Recherches sur la philosophie et le langage, n°23, Université Pierre Mendès France, Grenoble.

- CLERO, J.P. [2006], *Bentham: Philosophe de l'utilité*, Ellipses, Paris.
- CLERO, J.P. [2013], « Bentham et l'utilité », *Revue Internationale de Philosophie*, vol.263, n°1, p.99-112.
- CLERO, J.P. [2015], « Le calcul benthamien des plaisirs et des peines. Calcul introuvable ou seulement indéfiniment différé ? », *Archives de Philosophie*, vol.78, n°2, p.229-258.
- COT, A. [1992], "Jeremy Bentham, un 'Newton de la morale'", in A. BERAUD et G. FACCARELLO (eds.), *Nouvelle histoire de la pensée économique. Des scolastiques aux classiques*, vol. 1, La Découverte, Paris, p289-301.
- COT, A.L. [1996], « Une utopie utilitariste : Jeremy Bentham et les colonies », *Cahiers d'Economie Politique*, n°27-28 p. 193-210.
- CRIMMINS, J. [2014], "Bentham and utilitarianism in the early nineteenth century", in B. EGGLESTON & D. MILLER (Eds.), *The Cambridge Companion to Utilitarianism*, Cambridge University Press, Cambridge, p. 38-60.
- CRISP, R. [1997], *Mill on Utilitarianism*, Routledge, London.
- CRISP, R. [2002], *Routledge Philosophy GuideBook to Mill on Utilitarianism*, Taylor & Francis, Londres.
- CRISP, R. [2006], "Hedonism Reconsidered", *Philosophy and Phenomenological Research*, Vol. LXXIII, n°3, p.619-645.
- CRISP, R. [2014], "Sidgwick and utilitarianism in the late nineteenth century", in B. EGGLESTON & D. MILLER (Eds.), *The Cambridge Companion to Utilitarianism*, Cambridge University Press, Cambridge, p. 81-102.
- CRISP, R. [2015], *The Cosmos of Duty. Henry Sidgwick's Methods of Ethics*, Oxford University Press, Oxford.
- DANCY, J. [2006], 'Nonnaturalism', in Coop, D. (ed.), *The Oxford Handbook of Ethical Theory*, Oxford University Press, New-York, p.122-145.
- DARWALL, S.L. [1974], "Pleasure as Ultimate Good in Sidgwick's Ethics", *The Monist*, vol.58, n°3, p.475-489.
- DE CHAMPS, E. [2008], *"La déontologie politique", ou, La pensée constitutionnelle de Jeremy Bentham*, Librairie Dalloz, Genève.
- DE CHAMPS, E. et CLERO, J.P. (eds.) [2009], *Bentham et la France: Fortune et infortunes de l'utilitarisme*, numéro spécial, *Studies on Voltaire and the Eighteenth Century*, vol.9.
- DE CHAMPS, E. [2015], *Enlightenment and Utility. Bentham in French, Bentham in France*, Cambridge University Press, Cambridge.
- DE LAZARI-RADEK, K. et SINGER P. [2014], *The Point of View of the Universe: Sidgwick & Contemporary Ethics*, Oxford University Press, Oxford.
- DE LAZARI-RADEK, K. et SINGER P. [2017], *Utilitarianism. A Very Short Introduction*, Oxford University Press, Oxford.
- DELON, N. [2016], « Un Singer peut-il en remplacer un autre ? », *Klesis*, n°32, p.150-190.
- DONNER, W. [1998], "Mill's Utilitarianism", in Skorupski, J. (ed.), *The Cambridge Companion to Mill*, Cambridge University Press, Cambridge, p.255-292.
- DRIVER, J. [2014a], "Global Utilitarianism", in B. EGGLESTON & D. MILLER (Eds.), *The Cambridge Companion to Utilitarianism*, Cambridge University Press, Cambridge, p.166-176.

- DRIVER, J. [2014b], "The History of Utilitarianism", in ZALTA, E.N. (ed.), *The Stanford Encyclopedia of Philosophy*,  
<<https://plato.stanford.edu/archives/win2014/entries/utilitarianism-history/>>
- DUPUY, J.P. [1995], "Postface", in HALEVY E., *La formation du radicalisme philosophique*, vol.1 *La jeunesse de Bentham 1776-1789*, Presses Universitaires de France, Paris, p. 329-359.
- DWORKIN, R. [1977] *Prendre les droits au sérieux*, trad. Marie-Jeanne Rossignol et Frédéric Limare, [1995], PUF, Paris.
- DWORKIN, R. [1984], "Rights as Trumps", in J.WALDRON (ed.), *Theories of Rights*, Oxford University Press, Oxford, p. 153-167.
- ELSTER, J., [1983], *Sour Grapes: Studies in the Subversion of Rationality*, Cambridge University Press, Cambridge.
- FELDMAN, F. [2004], *Pleasure and the Good Life*, Oxford University Press, New York-Oxford.
- FLEURBAEY, M. et MONGIN, P. [1996], "Théorie du choix social et économie normative", in CANTO-SPERBER M., *Dictionnaire de philosophie morale*, Presses Universitaires de France, Paris, p. 243-51.
- FRANKENA, W. K. 1974. "Sidgwick and the dualism of practical reason", *The Monist*, vol.58, n°3, p.449-467.
- FRANKENA, W.K. [1992], "Sidgwick and the history of ethical dualism", in Schultz, B. *Essays on Henry Sidgwick*, Cambridge University Press, p.175-198.
- FOUCAULT, M. [1975], *Surveiller et punir. Naissance de la prison*, Gallimard, Paris.
- GALBRAITH, J.K., [1979], *The nature of mass poverty*, Harvard University Press, Cambridge, Mass. (USA).
- GILARDONE, M. [2007], *Contexte, sens et portée de l'approche par les capacités de Amartya Kumar Sen: Vers une économie normative post-welfariste*, Thèse de doctorat, Université Lyon 2, Lyon.
- GLOVER, J. (ed.) [1990], *Utilitarianism and its Critics*, Macmillan Publishing Company, New York-Londres.
- GLOVER, J. [2017], *Questions de vie ou de mort*, traduction et introduction de Benoît Basse, Labor et Fides, Genève.
- GODWIN, W. [1793], *An Enquiry Concerning Political Justice, and its Influence on General Virtue and Happiness*, G.G.J. and J. Robinson, London.
- GRAY, J. [1991], "Mill's Conception of Happiness and the Theory of Individuality", in Gray, J. et Smith, G.W. (eds.), *J.S. Mill On Liberty in focus*, Routledge, London and New York, p.190-211.
- GRIFFIN, J. [1986], *Well-being: its meaning, measurement, and moral importance*, Clarendon Press, Oxford.
- GRUEN, L. [1999 ], "Must Utilitarians Be Impartial?" in JAMIESON, D., *Singer and his Critics*, Blackwell Publishing, Oxford, p.129-149.
- GRUEN, L. [2016], "Utiliser la Philosophie pour Changer le Monde", *Klesis*, n°32, p.15-39.
- GUIDI, M.E.L. [2008], « 'Everybody to count for one, nobody for more than one': The Principle of Equal Consideration of Interests from Bentham to Pigou", *Revue d'études benthamiennes*, n°4, en ligne.
- GUILLOT, A. [2011], "L'utilitarisme Classique et le sujet de l'égalité », in BOZZO-REY, M. et DARDENNE, E., *Deux siècles d'utilitarisme*, Presses Universitaires de Rennes, Rennes, p. 119-129.



- GUILLOT, A. [2014], *Jeremy Bentham : le peuple comme fiction*, Publications de la Sorbonne, Paris.
- GUSTAFSON A. [2013], "In Defense of a Utilitarian Business Ethic", *Business and Society Review*, vol.118, n°3, p.325-360.
- HALEVY, E. [1901-1904], *La formation du radicalisme philosophique*, 3 vols., réed. [1995], Presses universitaires de France, Paris.
- HAMPSHIRE, S. [1972], *Freedom of mind, and other essays*, Clarendon Press, Oxford.
- HARE, R.M. [1981], *Moral Thinking: Its Levels, Method, and Point*, Oxford University Press, Oxford.
- HARE, R.M., [1982], "Ethical theory and utilitarianism", in SEN, A. et WILLIAMS, B. (eds.), *Utilitarianism and beyond*, Cambridge University Press, Cambridge, p.23-38.
- HARE, R.M., [1989], *Essays on Political Morality*, Oxford University Press, Oxford.
- HARE, R.M., [1999], "Why I Am Only a Demi-vegetarian", in JAMIESON, D., *Singer and his Critics*, Blackwell Publishing, Oxford, p.233-246.
- HARRISON, R. [1983], *Bentham*, Routledge & Kegan Paul, Londres.
- HARROD, R. F. [1936], "Utilitarianism revised", *Mind*, vol.45, n°178, p.137-156.
- HARSANYI, J.C. [1955], "Cardinal welfare, individualistic ethics and interpersonal comparisons of utility", *Journal of Political Economy*, vol.63, n°4, p.309-21.
- HARSANYI, J. C. [1982], "Morality and the Theory of Rational Behavior", in SEN, A. et WILLIAMS, B. (eds.), *Utilitarianism and beyond*, Cambridge University Press, Cambridge, p.39-62.
- HARSANYI, J. C. [1995], "A theory of prudential values and a rule utilitarian theory of morality", *Social Choice and Welfare*, vol. 12, n°4, p. 319-333.
- HASLETT, D.W. [1990], "What is utility?", *Economics and Philosophy*, vol.6, n°1, p.65-94.
- HAYBRON, D. M. [2016], "Mental State Approaches to Well-Being" In M.D. Adler et M. Fleurbaey (eds.), *The Oxford Handbook of Well-Being and Public Policy*, Oxford University Press, Oxford, pp. 347-378.
- HAYEK, F.A. [1969], *John Stuart Mill and Harriet Taylor: Their Correspondence and Subsequent Marriage*, University of Chicago Press, Chicago.
- HODGSON, D.H. [1967], *Consequences of Utilitarianism*, Clarendon Press, Oxford.
- HOOKE B. [1990], "Rule-Consequentialism", *Mind*, vol. 99, n°393, p.67-77.
- HOOKE, B. [2000], *Ideal Code, Real World: A Rule-consequentialist Theory of Morality*, Oxford University Press, Oxford.
- HOOKE, B. [2016], "Rule Consequentialism", in ZALTA E.N., *The Stanford Encyclopedia of Philosophy*,  
<<https://plato.stanford.edu/archives/win2016/entries/consequentialism-rule/>>
- HURKA, T. [1990], "Two Kinds of Satisficing", *Philosophical Studies*, vol.59, p.107-111.
- HURKA, T. [2003], "Moore in the Middle", *Ethics*, vol.113, n°3, p.599-628.
- HUTCHINSON, B. [2001], *G. E. Moore's Ethical Theory: Resistance and Reconciliation*, Cambridge University Press, Cambridge.
- JAQUET, F. [2016], "Utilitarisme", version académique, dans KRISTANEK, M. (ed.), *l'Encyclopédie philosophique*, URL: <http://encyclo-phil.fr/utilitarisme/>
- KAGAN, S. [2019], *How to Count Animals, more or less*, Oxford University Press, Oxford.
- KAHNEMAN, D., et TVERSKY A. [1979], "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, vol.47, n°2, p.263-291.
- KAHNEMAN, D., WAKKER, P.P. et SARIN R. [1997], "Back to Bentham? Explorations of Experienced Utility", *The Quarterly Journal of Economics*, vol.112, n°2, p.375-406.

- KAMM, F.M. [1999], "Faminine Ethics: the Problem of Distance in Morality and Singer's Ethical Theory", in JAMIESON, D., *Singer and his Critics*, Blackwell Publishing, Oxford, p.162-208.
- KANDIL, F. [2014], "La justice est aveugle: Rawls, Harsanyi et le voile d'ignorance", *Revue économique* vol.65, n°1, p.97-124.
- KAPLOW, L. [2007], "Primary Goods, Capabilities,... or Well-Being?", *The Philosophical Review*, vol.116, n°4, p.603-632.
- KNISS, J. [2019], "Bentham on animal welfare", *British Journal for the History of Philosophy*, vol.27, n°3, p.556-572,
- KYMLICKA, W., [2003], *Les théories de la justice: Une introduction: Libéraux, utilitaristes, libertariens, marxistes, communautariens, féministes..*, La Découverte, Paris.
- LAVAL, C. [1994], *Jeremy Bentham : Le pouvoir des fictions*, Presses Universitaires de France, Paris.
- LAVAL, C. [2003], *Jeremy Bentham, les artifices du capitalisme*, Presses Universitaires de France, Paris.
- LEVY, P. [1979], *Moore: G.E. Moore and the Cambridge Apostles*, Holt, Rinehart and Winston, New York.
- LYONS, D. [1965], *Forms and Limits of Utilitarianism*, Clarendon Press, Oxford.
- LYONS, D. [1980], "Utility as a possible ground of rights", *Noûs*, vol.14, n°1, p.17-28.
- LYONS, D. [1991], *In the interest of the governed: a study in Bentham's philosophy of utility and law*, Oxford University Press, Oxford.
- LYONS, D. [1994], *Rights, Welfare, and Mill's Moral Theory*, Oxford University Press, New York.
- MACASKILL, W. [2015], *Doing Good Better: Effective Altruism and a Radical Way to Make a Difference*, Guardian Faber, London.
- MACKIE, J. L. [1976], "Sidgwick's Pessimism", *The Philosophical Quarterly*, vol.26, n°105, p.317-327.
- MARTINOIA, R. [2011], « 'Une triste fin pour un si grand travail' ? La révision de l'utilitarisme par Henry Sidgwick », *Oecononia. History/Methodology/Philosophy*, vol.1, n°2, p.171-193.
- MCGINN, C. [1999], "Our Duties to Animals and the Poor", in JAMIESON, D., *Singer and his Critics*, Blackwell Publishing, Oxford, p.150-161.
- STUART MILL, J. [1965-1991], *Collected Works of John Stuart Mill*, ed. J. Robson, University of Toronto Press, Toronto, 33 vols.
- MILLER, R. B. [2009], "Actual Rule Utilitarianism", *The Journal of Philosophy*, Vol. 106, No. 1, p. 5-28.
- MITCHELL, W. [1918], "Bentham's Felicific Calculus", *Political Science Quarterly*, vol.33, n°2, p. 161-183.
- MONGIN, P. [1994], "Harsanyi's Aggregation Theorem: Multi-Profile Version and Unsettled Questions", *Social Choice and Welfare* 11, pp.331-354.
- MONGIN, P. [1995], "L'utilitarisme originel et le développement de la théorie économique", Postface de HALEVY E. *Le radicalisme philosophique*, vol.3, Presses Universitaires de France, Paris, pp. 369-394.
- MOORE, G.E. [1912], *Ethics*, Williams and Norgate, Londres.
- MOORE, G.E. [1903], *Principia Ethica*, Revised Edition [1993], Cambridge University Press, Cambridge.
- MULGAN, T. [1993], "Slote's Satisficing Consequentialism", *Ratio*, vol. VI, n°2, p.121-134.

- MULGAN, T. [2001], *The Demands of Consequentialism*, Clarendon Press, Oxford.
- MULGAN, T. [2006], *Future People. A Moderate Consequentialism Account of our Obligations to Future Generations*, Clarendon Press, Oxford.
- MULGAN, T. [2011], *Ethics for a Broken World. Imagining Philosophy After Catastrophe*, Acumen, Durham.
- MULGAN, T. [2018], "Answering to future people", *Journal of Applied Philosophy*, vol.35, n°3, p.532-48.
- MULGAN, T. [2020], *Utilitarianism*, Cambridge University Press, Cambridge.
- NAGEL, T. [1970], *The Possibility of Altruism*, Clarendon Press, Oxford.
- NAGEL, T. [1973], "Rawls on Justice", *The Philosophical Review*, vol. 82, n°2, p.220-234.
- NOZICK, R., [1974], *Anarchy, State, and Utopia*, traduction française [1988], Presses Universitaires de France, Paris .
- O'CONNOR, D. [1982], *The Metaphysics of G.E. Moore*, D. Reidel Publishing Company, Dordrecht.
- OLSON, E.T. [1997], *The Human Animal: Personal Identity Without Psychology*, Oxford University Press, Oxford.
- ORAZI, F. [2018], *L'individu libre. Le libéralisme anglo-saxon de John Stuart Mill à nos jours*, Classiques Garnier, Paris.
- ORSI, F. [2008], "The Dualism of the Practical Reason: Some Interpretations and Responses", *Etica & Politica / Ethics & Politics*, vol.X, n°2, pp. 19-41.
- OTSUKA, M. [2015], "Prioritarianism and the measure of utility", *Journal of Political Philosophy*, vol.23, n°1, p.1-22.
- OTSUKA, M et VOORHOEVE A. [2009], "Why it Matters that some are worse off than others: An Argument against the priority view", *Philosophy & Public Affairs*, Vol. 37, n° 2, p. 171-199.
- PARIENTE-BUTTERLIN, I. [2016], "Singer et la question de l'omission (d'un point de vue méta-éthique)", *Klesis*, vol.32, p.191-204.
- PARFIT, D. [1984], *Reasons and Persons*, Clarendon Press, Oxford.
- PARFIT, D. [1986], "Overpopulation and the Quality of Life", in SINGER, P. (ed.), *Applied Ethics*, Oxford University Press, Oxford, p.145–164.
- PARFIT, D. [1997], "Equality and Priority", *Ratio*, vol.10, n°3, p.202-21.
- PARFIT, D. [2011], *On What Matters*, vols. 1 et 2, Oxford University Press, Oxford.
- PARFIT, D. [2017], *On What Matters*, vol.3, Oxford University Press, Oxford.
- PETTIT P. et BRENNAN G. [1986], "Restrictive Consequentialism", *Australasian Journal of Philosophy*, Vol. 64, n° 4, p.438-455.
- PETTIT P. [1988], "The Consequentialist Can Recognise Rights", *The Philosophical Quarterly*, Vol. 38, n°150, p. 42-55.
- PETTIT, P. [1997], *Républicanisme. Une théorie de la liberté et du gouvernement*, traduit en [2004], Gallimard, Paris.
- PHILLIPS, D. [2011], *Sidgwickian Ethics*, Oxford University Press, Oxford.
- POGGE, T. (2002), *World Poverty and Human Rights*, Polity Press, Cambridge.
- POPPER, K. [1952], *The Open Society and Its Enemies*, réédition [2003], vol.1, Routledge, Londres.
- POSTEMA, G.J. [2006], "Interests, Universal and Particular: Bentham's Utilitarian Theory of Value", *Utilitas* vol.18, n°2, p. 109–33.
- POSTEMA, G.J. [2019], *Utility, Publicity, and Law: Essays on Bentham's Moral and Legal Philosophy*, Oxford University Press, Oxford.

- QUINN, M. [2014], "Bentham on Mensuration: Calculation and Moral Reasoning", *Utilitas*, vol.26, n°1, p.61-104.
- RAILTON, P. [1984], "Alienation, consequentialism, and the demands of morality", *Philosophy and Public Affairs*, vol.13, n°2, p.134-171.
- RASHDALL, H. [1907], *The Theory of Good and Evil*, Clarendon Press, Oxford.
- RAWLS, J. [1955], "Two Concepts of Rules", *Philosophical Review*, vol.64, n°1, p.3-32.
- RAWLS, J. [1971]. *Théorie de la justice*, traduit par C. AUDARD, [1987], Editions du Seuil, Paris.
- REGAN, D. [2003], "How to Be a Moorean", *Ethics*, vol.113, n°3, p.651-677.
- REGAN, T. [2012], *Bloombsury's Prophet: G. E. Moore and the Development of His Moral Philosophy*, Wipf & Stock Publishers, Eugene (Oregon).
- RILEY, J. [1998], *Routledge Philosophy Guidebook to Mill on Liberty*, Routledge, Oxon.
- RILEY, J. [1999], 'Is Qualitative Hedonism Incoherent?', *Utilitas*, vol.11, n°3, p.347-58.
- RILEY, J. [2003], 'Interpreting Mill's Qualitative Hedonism', *The Philosophical Quarterly*, vol.53, n°210, p.410-18.
- RISSE, M. [2002], "Harsanyi's 'Utilitarian Theorem' and Utilitarianism", *Noûs*, vol.36, n°4, p.550-577.
- ROEMER, J. E. [2008], "Harsanyi's impartial observer is not a utilitarian." In FLEURBAEY M., SALLES M. et WEYMARK J., *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*, Cambridge University Press, Cambridge, p. 129-135.
- ROSEN, F. [2003], *Classical Utilitarianism From Hume to Mill*, Routledge, Londres.
- ROSS, W.D. [1930], *The Right and the Good*, Oxford University Press, Oxford.
- RÜEGGER, M. [2011], "L'utilitarisme et le problème des préférences externes", in BOZZO-REY, M. et DARDENNE, E. (eds.), *Deux siècles d'utilitarisme*, Presses Universitaires de Rennes, Rennes.
- RYAN, A. [1974], *J. S. Mill*, Routledge and Kegan Paul, Londres.
- RYAN, A. [1998], "Mill in a Liberal Landscape", in Skorupski, J. (ed.), *The Cambridge Companion to Mill*, Cambridge University Press, Cambridge, p.497-540.
- SALVAT C., [2015], "Parfit, l'égoïsme rationnel et la question de l'identité personnelle", *Oeconomia*, vol.5, n°4, p.437-460.
- SALVAT, C. [2019a], « Parfit », version académique, dans Kristanek, M. (ed.), *l'Encyclopédie philosophique*, URL : <http://encyclo-philos.fr/parfit-a-2/>
- SALVAT C., [2019b], "Parfit et les raisons du conséquentialisme de la règle", *Revue de Métaphysique et de Morale*, n°2, p.39-50.
- SAMUELSON, P.A. [1938], "A Note on the Pure Theory of Consumer Behavior", *Economica*, vol. 5, n°17, p.61-71.
- SAUNDERS, B. [2016 ], "Recent Critics of Mill's Qualitative Hedonism", *Philosophy*, vol.91, n°4, p.503-521.
- SCANLON, T.M. [1982], "Contractualism and Utilitarianism", in SEN, A. et WILLIAMS, B. (eds.), *Utilitarianism and beyond*, Cambridge University Press, Cambridge, p.103-128.
- SCHEFFLER, S. [1982], *The Rejection of Consequentialism: a Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*, Clarendon Press-Oxford University Press, Oxford-New York.
- SCHOFIELD, P. [2006], *Utility and Democracy. The Political Thought of Jeremy Bentham*, Oxford University Press, Oxford.
- SCHOFIELD, P. [2009], *Bentham: A Guide for the Perplexed*, , Continuum, London.

- SCHOFIELD, P. [2014], "Jeremy Bentham on Taste, Sex, and Religion", in X. ZHAI and M. QUINN (eds), *Bentham's Theory of Law and Public Opinion*, Cambridge University Press, Cambridge, p. 90-118.
- SCHULTZ, B. [2004], *Henry Sidgwick - Eye of the Universe: An Intellectual Biography*, Cambridge University Press, Cambridge.
- SCHULTZ, B. [2007], "Mill and Sidgwick, Imperialism and Racism", *Utilitas*, vol.19, n°1, p. 104-130.
- SCHULTZ, B. [2017], *The Happiness Philosophers: The Lives and Works of the Great Utilitarians*, Princeton University Press, Princeton, (N.J., USA).
- SCHULTZ, B. et VAROUXAKIS, G. (eds.) [2005], *Utilitarianism and Empire*, Lexington Books, Lanham.
- SCHNEEWIND, J. B. [1977], *Sidgwick's Ethics and Victorian Moral Philosophy*, Clarendon Press, Oxford.
- SEN, A.K. [1977], "Non-linear Social Welfare Functions: A Reply to Professor Harsanyi." In BUTTS R. E. and HINTIKKA J. (eds), *Foundational Problems in the Special Sciences*, Springer, Dordrecht, p. 297–302.
- SEN, A.K. [1970], "The Impossibility of a Paretian Liberal", *Journal of Political Economy*, vol.78, n°1, pp.152–157.
- SEN, A.K. [1979], "Utilitarianism and Welfarism", *Journal of Philosophy*, vol.76, n°9, p.463–488.
- SEN, A.K. [1980], "Equality of What?" In S. McMurrin (ed.), *Tanner Lectures on Human Values, Volume 1*, Cambridge University Press, Cambridge.
- SEN, A.K. [1982], "Rights and Agency", *Philosophy & Public Affairs*, vol. 11, n° 1, p. 3-39.
- SEN, A. et WILLIAMS, B. (eds.) [1982], *Utilitarianism and beyond*, Cambridge University Press, Cambridge.
- SHAW, W. [1999], *Contemporary Ethics: Taking Account of Utilitarianism*, Wiley-Blackwell, Londres
- SHAW, W.H. [2016], *Utilitarianism and the Ethics of War*, Routledge, Londres.
- SHOEMAKER S. et SWINBURNE R. [1984], *Personal Identity*, Blackwell, Oxford.
- SHOEMAKER S. [1997], "Self and Substance", *Noûs*, vol. 31, n°11, p.283-304.
- SIDGWICK, H. [1874], *The Methods of Ethics*, seventh reedition [1981] Hackett Publishing Company, Indianapolis.
- SIGOT, N. [2011], *Bentham et l'économie. Une histoire de l'utilité*, Economica, Paris.
- SIMON, H. [1955], "A Behavioral Model of Rational Choice", *Quarterly Journal of Economics*, vol.69, n°1, p.99-118.
- SINGER, M.G. [1992], "Sidgwick and nineteenth-century British ethical thought", in Schultz, B. (ed.), *Essays on Henry Sidgwick*, Cambridge University Press, Cambridge, p. 65-91.
- SINGER, P. [1972], 'Famine, Affluence, and Morality', *Philosophy & Public Affairs*, vol.1, n°3, p.229-43.
- SINGER, P. [1975], *La libération animale*, traduit pat L. ROUSSELLE, [2012], Petite Bibliothèque Payot, Paris.
- SINGER, P. [1979], *Practical Ethics*, Cambridge University Press, Cambridge.
- SINGER, P. [1998], *Théorie du tube de dentifrice. Comment changer le monde selon Henry Spira*, traduction par A. PONS, [2018], Editions Goutte d'Or, Paris.
- SINGER, P. [2015], *The Most Good You Can Do: How Effective Altruism Is Changing Ideas About Living Ethically*, Yale University Press, New Haven and London.

- SINGER, P. [1997], *Questions d'éthique pratique*, trad. Marcuzzi M., Bayard Editions, Paris.
- SINNOTT-ARMSTRONG, W., [2019], "Consequentialism", in ZALTA, E.N., *The Stanford Encyclopedia of Philosophy*,  
 <<https://plato.stanford.edu/archives/sum2019/entries/consequentialism/>>
- SKELTON, A. [2011], "Ideal Utilitarianism: Rashdall and Moore." In T. Hurka (ed.), *Underivative Duty: British Moral Philosophers from Sidgwick to Ewing*, Oxford University Press, Oxford, p. 45–65.
- SKORUPSKI, J. [1989], *John Stuart Mill*, Routledge & Kegan Paul, Londres.
- SKORUPSKI, J. [2001], "Three Methods and a Dualism", *Proceedings of the British Academy*, vol.109, p.61-81.
- SLOTE, M.A. [1983], *Goods and Virtues*, Clarendon Press, Oxford.
- SLOTE, M.A. [1985], *Common-Sense Morality and Consequentialism*, Routledge & Kegan Paul, Londres.
- SLOTE, M.A., et PETTIT P. [1984], "Satisficing Consequentialism", *Proceedings of the Aristotelian Society, Supplementary Volumes* 58, p.139-176.
- SMART, J.J.C. [1956], "Extreme and restricted utilitarianism", *Philosophical Quarterly*, vol.6, n°25, p.344-354.
- SMART, J.J.C., et WILLIAMS B. [1973], *Utilitarianism: For and Against*, Cambridge University Press, Cambridge.
- SMART, J.J.C. [1978], J. J. C. Smart, "Utilitarianism and Justice," *Journal of Chinese Philosophy*, vol. 5 n°3, p.287-299; reprinted in GLOVER J. [1990], *Utilitarianism and its Critics*, Macmillan Publishing Company, New York, p.170-174.
- SMART, R. N. [1958], "Negative Utilitarianism", *Mind*, vol.67, n°268, p.542-543.
- SPITZ, J.-F. [2013], "Le conséquentialisme hybride de Samuel Scheffler : défense et illustration", in S. GUERARD DE LATOUR et M.-A. DILHAC (eds.), *Étant donné le pluralisme*, Éditions de la Sorbonne, Paris, p.27-53.
- SUMNER, L. W. [1996], *Welfare, Happiness, and Ethics*, Oxford University Press, New York-Oxford.
- TÄNNSJÖ, T. [1976], *The Relevance of Metaethics to Ethics*, Almqvist & Wiksell International, Stockholm.
- TÄNNSJÖ, T. [2002], "Why We Ought to Accept the Repugnant Conclusion", *Utilitas*, vol.14, n°3, p.339 - 359.
- TEMKIN, L. S. [1987], "Intransitivity and the Mere Addition Paradox", *Philosophy and Public Affairs*, vol.16, n°2, p.138–187.
- TEN, C.L. [1980], *Mill on Liberty*, Clarendon Press, Oxford.
- TERESTCHENKO, M. [2004], "Henry Sidgwick: Le cosmos de la moralité réduit au chaos », *Revue de Métaphysique et de morale*, vol.41, n°1, p.101-128.
- THOMAS, W. [1993], 'L'utilitarisme et le libéralisme anglais au début du XIXe siècle', in MULLIGAN, K. and ROTH, R. (eds.) *Regards sur Bentham et l'utilitarisme*, Librairie Droz, Genève.
- TOULMIN, S.E. [1950], *An Examination of the Place of Reason in Ethics*, Cambridge University Press, Cambridge.
- TUCKER, C. [2017], "How to Think about Satisficing", *Philosophical Studies*, vol.174, n°6, p.1365-1384.
- TUSSEAU G. (ed.) [2014], *The Legal Philosophy and Influence of Jeremy Bentham: Essays on 'Of the Limits of the Penal Branch of Jurisprudence'*, Routledge, Oxon.
- TUSSEAU, G. [2014], *Jeremy Bentham. La guerre des mots*, Librairie Dalloz, Genève.

- TUSSEAU, G. (ed.) [2019], *La déontologie publique : trajectoire et présence d'une notion ambiguë*, Institut Francophone pour la Justice et la Démocratie, Paris.
- URMSON, J. O. [1953], "The Interpretation of the Moral Philosophy of J. S. Mill", *The Philosophical Quarterly*, vol. 3, n°10, p.33-39.
- WARKE, T. [2000], "Multi-Dimensional Utility and the Index Number Problem: Jeremy Bentham, J.S. Mill and Qualitative Hedonism", *Utilitas*, vol. 12, n°2, p.176–203.
- WEST, H. [2007], *Mill's Utilitarianism*, Continuum, London.
- WEST, H. [2014], "Mill and utilitarianism in the mid-nineteenth century", in B. EGGLESTON & D. MILLER (Eds.), *The Cambridge Companion to Utilitarianism*, Cambridge University Press, Cambridge, p. 61-80.
- WEYMARK, J. A. [1991], "A Reconsideration of the Harsanyi–Sen Debate on Utilitarianism." In ELSTER J. and ROEMER J. E (eds.), *Interpersonal Comparisons of Well-being*, Cambridge University Press, Cambridge, p. 255–320.
- WHATMORE, R. [2007], "Etienne Dumont, the British Constitution, and the French Revolution", *Historical Journal*, vol.50, n°1, p. 23-47.
- WILLIAMS, B. [1973], "A Critique of Utilitarianism", in SMART J.J.C et WILLIAMS B., *Utilitarianism: For and Against*, Cambridge University Press, Cambridge, p.75-150.
- WILLIAMS, B. [1973], "Persons, Character, and Morality", in *Moral Luck: Philosophical Papers 1973–1980*, Cambridge University Press, Cambridge, p.1-19.
- ZHAI, X. et QUINN, M. (eds.) [2014], *Bentham's Theory of Law and Public Opinion*, Cambridge University Press, Cambridge.
- ZUBER S. [2019], " Parfit et les économistes : la contribution de Derek Parfit au débat sur la taille optimale de la population ", *Revue de Métaphysique et de Morale*, n°2, p.23-37.

## Sélection de travaux (II)

### Articles et chapitres d'ouvrages

- "Still-Born Yet not Without Influence: What Mill's Political Economy Owes to His Project of Ethology", *Journal of the History of Economic Thought*, à paraître (2021).
- "Parfit et les raisons du conséquentialisme de la règle", *Revue de Métaphysique et de Morale*, n°2, 2019, pp.37-48.
- "L'éthique face à la GPA", *Raison Publique*, n°23, 2019, pp.163-73.
- "Parfit l'égoïsme rationnel et la question de l'identité personnelle", *Oeconomia, History, Methodology, Philosophy*, 5(4), 2015, pp.437-60.
- "Economics of paternalism: the hidden costs of self-commanding strategies", *The Journal of Philosophical Economics: Reflections on Economic and Social Issues*, 9(1), 2015, pp.102-124.
- "Behavioral Paternalism", *Revue de Philosophie Economique*, vol.15, n°2, 2014, pp.109-130.
- "Rousseau et la « Renaissance classique » française (1898-1933) ", *Astérion, Philosophie, Histoire des Idées, Pensée politique*, 12, 2014  
[<http://asterion.revues.org/2545>]
- "Autonomie morale et autorité, Ou la question de la volonté chez Rousseau", *Cahiers d'Economie Politique*, n°53, Décembre 2007, 75-93
- "La loi et l'échange: le statut de la rationalité économique chez Rousseau", *Revue économique*, volume 58, n°2, Mars 2007, pp.381-398.
- "Oeconomies. Les articles OE/Economie et leurs désignants dans l'Encyclopédie", *Recherches sur Diderot et l'Encyclopédie*, n°40, Octobre 2006, pp.97-116.





JOURNAL OF THE  
**HISTORY OF  
ECONOMIC THOUGHT**

PUBLISHED FOR THE HISTORY OF ECONOMICS SOCIETY

**Still-Born Yet Not Without Influence: What Mill's Political  
Economy Owes to his Project of Ethology**

Journal:	<i>Journal of the History of Economic Thought</i>
Manuscript ID	HET-18-09-Res-072.R3
Manuscript Type:	Research Article
Abstract:	

SCHOLARONE™  
Manuscripts

## STILL-BORN YET NOT WITHOUT INFLUENCE

## WHAT MILL'S POLITICAL ECONOMY OWES TO HIS PROJECT OF ETHOLOGY

By

CHRISTOPHE SALVAT

*Abstract:* This article questions the articulation between John Stuart Mill's initial project of creating a new science dedicated to the means of improving individual character, a science named ethology, and the treatise of political economy that he published instead. My claim is that his defence of free competition as well as some of the arguments he opposes to it, and which have often puzzled his readers, actually reveal the moral agenda of his political economy and of some of his political principles, specifically his ambivalent position towards paternalism.

## I. INTRODUCTION

In chapter five of book six of the *System of Logic*, John Stuart Mill proposed to found a new science, referred to as ethology. The aim of this new science would be to understand the role played by social circumstances in individual choices and to improve people's character by intervening in their social environment. This science, if it had been achieved, would have had – alongside political economy – a significant place in Mill's moral system. In its absence, political economy seems almost at odds with the general moral background of his philosophy. However,

1  
2  
3 the main consequence of its absence is to conceal the perfectionist or paternalistic feature of  
4 Mill's general framework. This paper therefore discusses the theoretical and practical  
5 implications of the missing theory of ethology relatively to political economy. It shows that,  
6 despite not being the object of a full treatise, ethology is nonetheless present in Mill's  
7 philosophical system and that it surreptitiously resurfaces, notably, in his political economy and  
8 his politics. It also highlights the importance given to excellence or nobleness of character, a  
9 notion too often discarded as a mere trace of the Victorian frame of mind (Houghton 1957).  
10 Failing to address the issue of character formation in Mill's thought could, I contend, lead to  
11 serious misunderstandings of his philosophy, particularly in relation to paternalism.  
12  
13 The first section presents Mill's initial project of setting a new science inherited from  
14 associationism and directed towards the improvement of individual characters. Section two  
15 questions the nature of the improvement considered and suggests that the desired qualities  
16 alluded to in the *System of Logic* match those of the 'noble characters' he extensively develops in  
17 *Utilitarianism*. Section three introduces the main liberal features of his *Principles on political*  
18 *economy* published instead of the expected treatise on ethology. Sections four and five further  
19 examine the link between political economy and ethology. Whilst section four focuses on how  
20 improving characters justifies some decisive exceptions to the general principle of free  
21 competition (colonies, education and land property), section five argues that Mill sees in the  
22 principle of competition, and therefore in the promotion of political economy, one of the most  
23 efficient ways to improve individual characters. The final section questions the paternalistic  
24 feature of Mill's philosophy.

## 55 II. THE SCIENCE OF ETHOLOGY

1  
2  
3 In chapter five of book six of the *System of Logic*, Mill proposed to found a new science, called  
4  
5 ethology, which aimed to understand the role played by social circumstances in individual  
6  
7 choices. Based on psychology's general laws, ethology was supposed to reconcile deductive and  
8  
9 universal laws of human nature with seemingly contrary empirical evidence. Its aim was to  
10  
11 explain why individuals behave differently in comparable circumstances whilst all being subject  
12  
13 to the same universal laws. In his classic *History of English Utilitarianism* (1902), Ernest Albee  
14  
15 maintained that Mill abandoned his project of developing a science of ethology soon after the  
16  
17 publication of the *System of Logic* to dedicate himself to political economy (Albee 1902). But,  
18  
19 according to David Leary, Mill's failure to complete his scheme can be partly explained by  
20  
21 ethology's own weaknesses. As a deductive science-to-be grounded on empirical psychology,  
22  
23 ethology was, he believes, doomed from the start (Leary 1982, p.155). This theory has since been  
24  
25 contested (Capaldi 1973, Leary 1982, Ball 2000). I claim here that, whatever Mill's reasons for  
26  
27 abandoning his project of writing a full treatise on ethology, he never abandoned the idea of  
28  
29 using social science, and political economy in particular, to improve individual characters.  
30  
31  
32  
33  
34  
35

36 Although ethology is only explicitly dealt with in *System of Logic*, character formation is a topic  
37  
38 Mill often refers to in his other works. This is especially true in *An Examination of Sir William*  
39  
40 *Hamilton's Philosophy* (1865), *Auguste Comte and Positivism* (1865), and his notes on James  
41  
42 Mill's *Analysis of the Phenomena of the Human Mind* (1869). Carlisle (1991) and Ball (2000)  
43  
44 have argued that Mill's autobiography could be read as a case study of ethology, an illustration  
45  
46 of how "his character was being formed (and to some degree deformed) by his father and  
47  
48 subsequently re-formed by himself" (Bal 2000, p.33). The issue of the formation of character is  
49  
50 also a prominent one in *Considerations on Representative Government* (1861) and *Subjection of*  
51  
52 *Women* (1869). I shall, however, essentially deal here with his writings on political economy.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 As a first approximation, ethology can be described as a branch of human psychology. Its  
4  
5 origins are to be found in the associationist theory defended by David Hume, David Hartley,  
6  
7 Alexander Bain, as well as James Mill's *Analysis of the Phenomena of the Human Mind* (1829)<sup>1</sup> .  
8  
9 In his *Autobiography*, Mill describes his father's 'fundamental doctrine' as "the formation of all  
10  
11 human character by circumstances, through the universal Principle of Association, and the  
12  
13 consequent unlimited possibility of improving the moral and intellectual condition of mankind  
14  
15 by education. Of all his doctrines none was more important than this, or needs more to be  
16  
17 insisted on" (Mill [1874] 1981, p.111). Those views were, however, not generally shared,  
18  
19 including by associationist theorists. Robert Owen, for instance, contested the role (self-  
20  
21 )education could potentially play in character formation (Tait 1971). Auguste Comte, Mill's  
22  
23 main opponent, was however not a disciple of associationism, and he considered it to be  
24  
25 scientifically flawed<sup>2</sup>.  
26  
27  
28  
29  
30  
31  
32  
33

---

34  
35 <sup>1</sup> Despite his young age, it is believed that John Stuart assisted his father during the redaction of  
36  
37 the *Analysis* (Carlisle 2010, pp.18-19).

38  
39 <sup>2</sup> For Comte, for something to be considered as scientific observation, it must satisfy two  
40  
41 conditions: observations must be veridical and they ought to presuppose a prior theory. Yet these  
42  
43 two conditions cannot be fulfilled with introspection (Wilson 1991, p.118). According to the  
44  
45 French philosopher, the principle of interior observation belongs to the metaphysico-theological  
46  
47 stage of philosophy. For Comte, the identity between the observing and the observed organ  
48  
49 makes the "principle of this so-called psychological method [...] entirely worthless" (Comte,  
50  
51 1988, 21). Grounding individual differences on physiological factors, Comte is known for having  
52  
53 justified social inequalities such as the inferior social position of women, which he attributed to  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Although they often disagree on fundamental issues, associationist philosophers all agree that  
4 human beings follow general and universal psychology laws and that the distinctness of  
5 individual behaviours can be explained by the great variety of circumstances they are subject to.  
6  
7 As a member of the Associationist School (to which Mill's father also belonged), Mill believed  
8 that individual actions are caused by a number of psychological associations, and that provided  
9 we have sufficient knowledge of the character of the person and of her environment, then her  
10 actions could be scientifically predicted, since "volitions do, in point of fact, follow determinate  
11 moral antecedents with the same uniformity, and (when we have sufficient knowledge of the  
12 circumstances) with the same certainty, as physical effects follow their physical causes" (Mill  
13 [1865] 1979, p.446).  
14  
15

16  
17 In its purest form, associationism excludes the idea of free will. For Mill, however, individuals  
18 can – to some extent – alter their character (Mill [1865] 1979, p.466). In that respect, Mill  
19 dissociates himself from Owen and his followers, who believed that individual character could  
20 only be moulded through external action (Harrison 2009, Owen 1813). Determinism is not  
21 'Fatalism', says Mill, and it certainly does not diminish an individual's personal responsibility in  
22 their actions. This distinction between irresistible and resistible causes of action is, as suggested  
23 by John Skorupski, fundamental to avoid determinism turning into fatalism (Skorupski 1989,  
24 p.252). It would be misleading, however, to conclude that individuals owe their capacity to resist  
25 motives of action to their moral autonomy. They only owe it to their character. People who have  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

---

the smaller size of their brain. This particular question was a major bone of contention between  
both philosophers who eventually fell out in 1847 (Olivier 2002).

1  
2  
3 a strong character are better equipped to select amongst different motives of action the one they  
4 will actually yield to. But this is not autonomy, not – at least – in the Kantian meaning of the  
5 word<sup>3</sup>. To be fully autonomous, the agent's will ought to be independently and rationally formed.  
6  
7 For Mill, however, all motives of action are externally determined and nothing suggests that a  
8 rational – and consequently universal – one should be given priority. On the contrary, he clearly  
9 states in *On Liberty*, that all experiences of living are equally desirable and, moreover, that this  
10 includes the desires of the most spontaneous and impulsive people; the most praiseworthy, in  
11 effect, as they are the mark of a strong and energetic character.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21

22 The aim of ethology is “to deduce the requisite middle principles from the general laws of  
23 Psychology. The subject to be studied is, the origin and sources of all those qualities in human  
24 beings which are interesting to us, either as facts to be produced, to be avoided, or merely to be  
25 understood: and the object is, to determine, from the general laws of mind, combined with the  
26 general position of our species in the universe, what actual or possible combinations of  
27 circumstances are capable of promoting or of preventing the production of those qualities.” (Mill  
28 [1843] 1974, pp.873-74). If ethology can give us a scientific account of how our character, and  
29 as a consequence our behaviour, is determined by our environment, then it becomes possible to  
30 design policies that could positively affect people's decision-making processes, without directly  
31 interfering with their personal choices.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

---

54  
55 <sup>3</sup> On the different meanings autonomy can have in Mill's writings, see (Fuchs 2001).  
56  
57



1  
2  
3 There are two different sides to ethology depending on whether it applies to national or to  
4 individual character. Mill believed each developed nation had its particular character<sup>4</sup> and that it  
5 would be “an absurdity” (Mill [1865] 1979, p.99) to ignore these differences when discussing  
6 legal and political matters. This is, as John Robson showed it, the very core of Mill’s attack  
7 against Jeremy Bentham’s universalism. Similar points can be made against the supposed  
8 universality of the principles of political economy. Personal characters are the combined product  
9 of individual and social circumstances, and it is through a combined action on these  
10 circumstances that personal character can evolve. How could it be otherwise, knowing that we all  
11 influence our (close) environment, just as we are swayed by it? The real issue, though, is to what  
12 extent it is possible to anticipate the consequences on people’s character of a change in their  
13 environment; or, in other words, whether Mill’s project of a science of ethology is achievable.  
14 Mill believes that, if it is impossible to anticipate with great exactitude the consequences of a  
15 social policy in that respect, it is nonetheless possible – or at least it should soon be – to ascertain  
16 its general impact. This should suffice to decide whether or not an action is desirable: “It is  
17 enough that we know that certain means have a tendency to produce a given effect, and that  
18 others have a tendency to frustrate it. When the circumstances of an individual or of a nation are  
19 in any considerable degree under our control, we may, by our knowledge of tendencies, be  
20 enabled to shape those circumstances in a manner much more favourable to the ends we desire,  
21 than the shape which they would of themselves assume. This is the limit of our power; but within  
22 this limit the power is a most important one.” (Mill [1843] 1974, pp.869-870).

---

23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
<sup>4</sup> Mill believed, for instance, that the Germans were *speculative* while the British were *practical*.

For more details see (Robson 1998).

### III. THE ULTIMATE AIM OF ETHOLOGY

The cultivation of nobleness of character, for Mill, “should be to individual human beings an end” (Mill [1843] 1974, p.952), and therefore constitute the ultimate aim of ethology. The idea of nobleness of character is a recurrent notion in Mill’s writings, one that he adopted when still very much under the influence of the Romantics and one that he never abandoned<sup>5</sup>. Mill believed, as did a number of public moralists, that some types of characters, which he called ‘noble characters’, were of higher moral standards than others, and that the achievement of a well-ordered society depended on people’s ability to excel in character.

What does being of ‘noble character’ involve? Mill does not give a full account. Noble translates as love of pride, of liberty and personal independence, of power or excitement, “but its most appropriate appellation is a sense of dignity, which all humans beings possess in one form or other, and in some, though by no means in exact, proportion to their higher faculties, and which is so essential a part of the happiness of those in whom it is strong.” (Mill 1985, p.212). The notion of nobleness of character is purposefully kept ill-defined by Mill, who does not want to convey the idea that it only refers to a unique pattern. Noble characters are those able to assert their singularity, whilst at the same time cultivating higher pleasures and higher values.

---

<sup>5</sup> Mill commented that Bentham knew absolutely nothing about human character (Mill 1985, pp.92-93). Romantics, on the other hand, realized its importance but never approached it from a scientific point of view.

1  
2  
3 There are two main institutional ways to foster nobleness of character for Mill. The first one is to  
4 further democracy. This may be surprising as Mill clearly singled out democracy as a major  
5 cause of standardization of characters. It is important, however, to distinguish the social and the  
6 political meaning of the word democracy. Democracy, as a social movement towards equality,  
7 and for Mill, like for Alexis de Tocqueville before him, inevitably results in the downwards  
8 harmonization of people's lifestyles. Democracy, as a political system of representative  
9 government appears to be the only appropriate response to what Tocqueville referred as the  
10 'fated' gradual progress of equality. But democracy can also represent a serious threat to  
11 nobleness of characters. Strong and energetic characters are an asset to the proper functioning of  
12 representative government but, as Mill is well aware, they are not usually appreciated by people<sup>6</sup>.  
13 They are, therefore, less likely to be elected. The British election system exacerbates this  
14 'natural' tendency. "At present", regrets Mill, "by universal admission, it is becoming more and  
15 more difficult for anyone, who has only talents and character, to gain admission into the House  
16 of Commons. The only persons who can get elected are those who possess local influence, or  
17 make their way by lavish expenditure, or who, on the invitation of three or four tradesmen or  
18 attorneys, are sent down by one of the two great parties from their London clubs, as men whose  
19 votes the party can depend on under all circumstances" (Mill [1843] 1974, pp.455-56). The real

---

20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46 <sup>6</sup> "The commonplaces of moralists, and the general sympathies of mankind, are in favour of the  
47 passive type. Energetic characters may be admired, but the acquiescent and submissive are those  
48 which most men personally prefer. The passiveness of our neighbours increases our sense of  
49 security, and plays into the hands of our wilfulness. Passive characters, if we do not happen to  
50 need their activity, seem an obstruction the less in our own path" (Mill [1843] 1974, p.407).

1  
2  
3 danger, as Mill sees it, is not that noble characters are underrepresented – this is to be expected –  
4 but that they are *totally* excluded from representation. Once a few are elected, Mill confidently  
5  
6 assumes that they will necessarily be heard. Although only counting according to their number in  
7  
8 the actual voting, the instructed minority “would count for much more, in virtue of their  
9  
10 knowledge, and of the influence it would give them over the rest” (Mill [1843] 1974, p.460).  
11  
12  
13

14  
15 The second<sup>7</sup> institutional lever to alter people’s character in England (as well as anywhere else)  
16  
17 is to promote, to generalise, and to improve education. Although all men have the potential to  
18  
19 experience such pleasures, some people’s pleasures are limited on account of their lack of  
20  
21 education. According to associationism, sensations produce ideas, which in turn, when  
22  
23 associated to other’s ideas, produce complex mental states that enable us to derive pleasure from  
24  
25 situations or events we would otherwise have been indifferent to. Education gives us the means  
26  
27 to develop the necessary faculties that will be later instrumental to our happiness. But education  
28  
29 is wasted if it stops individuals from asserting their character, as Mill believes. Strength and  
30  
31 diversity of character is what education is about, and more importantly even, it is what  
32  
33 civilization is about. “Ethology”, writes Mill, “is the science which corresponds to the art of  
34  
35  
36  
37  
38  
39  
40  
41  
42

---

43 <sup>7</sup> Education is only “second” in the order of this discussion. In fact, education has priority over  
44  
45 democracy for Mill: “When society has not performed its duty, by rendering this amount of  
46  
47 instruction accessible to all, there is some hardship in the case, but is a hardship that ought to be  
48  
49 borne. If society has neglected to discharge two solemn obligations, the more important and  
50  
51 more fundamental of the two must be fulfilled first: universal teaching must precede universal  
52  
53 enfranchisement.” (Mill 1977, p.470).  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 education in the widest sense of the term, including the formation of national or collective  
4 character as well as individual” (Mill [1843] 1974, p.869).  
5  
6  
7

8 It is therefore one of the government’s main duties and one of the very few, if not the only, case  
9 where it has full legitimacy to interfere with its citizen’s private lives: “Education [...] is one of  
10 those things which it is admissible in principle that a government should provide for the people.  
11  
12  
13

14 The case is one to which the reasons of the non-interference principle do not necessarily or  
15 universally extend.” (Mill [1848] 1965, p.948). In aspects of practice, promoting education is  
16 achieved by setting up new schools across the country, however, the most important aspect is  
17 rather the underlying legal obligation made for parents to send their children to school at little or  
18 no cost (Mill [1848] 1965, p.949). Schools are also not required to be state-funded, so that  
19 parents have the freedom, as they ought, to send their children to private and religious schools.  
20  
21  
22 Furthermore, the government should have no authority to bind teachers to a single curriculum  
23 (Mill [1848] 1965, p.950). The only obligation to be made is to provide proper education to all  
24 children. This, Mill believes, is far from being the case in Victorian England, where “[t]he  
25 present wretched education, and wretched social arrangements” are responsible of the English’s  
26 lack of sophistication, their egoism and their unrepentant sexism, and these should be the first  
27 institutions to be reformed (Mill [1861] 1985, p.215).  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

44 For Mill, nobleness of character refers to “the conformity of his own character to his standard of  
45 excellence, without hope of good and fear of evil from other source than his own inward  
46 consciousness.” (Mill [1838] 1985, p.95). It relies on a deep conviction that ‘individuality should  
47 assert itself’ (Mill [1859] 1977, p.261), that different modes of life are worth living and that ‘free  
48 scope should be given to varieties of character, short of injury to others’ (Mill [1859] 1977,  
49 p.261). This belief is supported by different types of arguments. The first one is utilitarian in  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 nature: by letting people find their pleasure wherever they choose to find it (on the condition  
4 that no one else is hurt) one ensures their greatest happiness.  
5  
6  
7

8 Mill secondly sees the cultivation of strong characters as a defence against the standardization of  
9 characters entailed by the upcoming democracy. Strong and energetic characters are impervious  
10 to public opinion and custom. “The general average of mankind”, regrets Mill, “are not only  
11 moderate in intellect, but also moderate in inclinations: they have no tastes or wishes strong  
12 enough to incline them to do anything unusual, and they consequently do not understand those  
13 who have, and class all such with the wild and intemperate whom they are accustomed to look  
14 down upon.” (Mill [1859] 1977, p.271). But for this standard of excellence to be achieved,  
15 adequate circumstances need to be created. This is where sciences such as ethology and political  
16 economy should assist the legislator.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

#### 33 IV. MILL’S CONCEPTION OF ECONOMICS 34 35

36 Mill’s *Principles of political economy* stands as one of the most influential economic textbooks  
37 of the second half the nineteenth century. Initially published in 1848, it has been continually  
38 republished, was revised by Mill up until its seventh edition in 1871, and was extensively used at  
39 least until World War One. Mill’s contribution to economics has often been undermined and read  
40 only to be compared to Adam Smith’s and David Ricardo’s. It is true that historians of economic  
41 thought, who used to exclusively focus on a retrospective history of economic concepts, have not  
42 been very sympathetic to Mill’s works (Schumpeter 1954, Blaug 1985, Hollander 1985). The  
43 idea according to which Mill’s economic ideas lacked originality was, however, not shared by  
44 everyone. John Hicks considered him as “the most undervalued economist of the nineteenth  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 century” (Hollander 1985, p.ix), and George Stigler even claimed that he was “one of the most  
4 original economists in the history of the science” (Hollander 1985, p.xii).  
5  
6  
7

8 Unlike Ricardo, Mill does not see the decline of the rate of profits as an adverse process to arrest  
9 or to delay but rather as a necessary condition to reach the stationary state, which he sees as “a  
10 very considerable improvement on [their] present condition” (Mill [1848] 1965, p.754)<sup>8</sup>. Since  
11 acquisition of wealth is made virtually impossible in such a state, people can concentrate on a  
12 better distribution: “It is only in the backward countries of the world that increased production is  
13 still an important object: in those most advanced, what is economically needed is a better  
14 distribution, of which one indispensable means is a stricter restraint on population” (Mill [1848]  
15 1965, p.755). Like Thomas Malthus, he disapproved of the unsustainable growth of population  
16 caused by people’s imprudent behaviour, but does not contemplate colonial trade as a possible  
17 solution to what can only be regarded as a *moral* issue. Political economy, on the contrary, offers  
18 conclusive evidence that population growth and profit rates are closely linked. While “the  
19 countries in which the greatest prudence is manifested in the regulating of population, are often  
20 those in which capital increases less rapidly”, one can also observe that “[w]here there is an  
21 indefinite prospect of employment for increased numbers, there is apt to appear less necessity for  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

---

42  
43 <sup>8</sup> To be more specific, two different situations should be distinguished. If the stationary state is  
44 brought on by a rise in Ricardian rent, wages will be left at subsistence, which hardly qualifies as  
45 an ‘improvement’. But if the stationary state is generated through moral restraint (hence  
46 controlling population growth), workers will benefit from better life conditions. I am thankful to  
47 an anonymous referee for pointing out this important distinction to me. For more details, see  
48 (Persky 2016, p. 85).  
49  
50  
51  
52  
53  
54  
55  
56  
57

1  
2  
3 prudential restraint” (Mill [1848] 1965, p.753). The economic slowdown consecutive to the fall  
4  
5 of the profit rate should therefore, for Mill, be conducive to an improvement of people’s moral  
6  
7 constraint. According to his own words, Mill “cannot [...] regard the stationary state of capital  
8  
9 and wealth with the unaffected aversion so generally manifested towards it by political  
10  
11 economists of the old school” (Mill [1848] 1965, pp.753-54).  
12  
13  
14

15  
16 The difficulty in assessing Mill’s contribution to political economy from a strict economic point  
17  
18 of view becomes obvious once we realize that, for Mill, the science of political economy, cannot  
19  
20 be considered separately from the other social sciences, and most specifically from ethology.  
21  
22 Political Economy, should not be treated “as a thing by itself, but as a fragment of a greater  
23  
24 whole; a branch of Social Philosophy, so interlinked with all the other branches, that its  
25  
26 conclusions, even in its own peculiar province, are only true conditionally, subject to interference  
27  
28 and counteraction from causes not directed within its scope: while to the character of a practical  
29  
30 guide it has no pretension, apart from other classes of considerations.” (Mill [1874] 1981, p.236).  
31  
32  
33  
34 Whilst economic commentators are prone to acknowledge the distinction introduced by Mill  
35  
36 between science and art, they are not ready to admit that, in order to have any possible  
37  
38 application, political economy must be connected to other social sciences.  
39  
40

41  
42 Science “deals in facts, [art] in precepts. Science is a collection of truths; art, a body of rules, or  
43  
44 directions for conduct” (Mill 1967, p.312). As a science, political economy is logically strictly  
45  
46 restricted to the uncovering of universal economic laws. To identify such laws, it is necessary to  
47  
48 imagine an economic agent stripped from all that could potentially makes him belong to a certain  
49  
50 time or a certain culture. “Just in the same manner [that geometry presupposes an arbitrary  
51  
52 definition of a line]”, writes Mill in his *Essays on Some Unsettled Questions of Political*  
53  
54 *Economy* (1844), “does Political Economy presuppose an arbitrary definition of a man, as a  
55  
56  
57  
58  
59  
60



1  
2  
3 being who invariably does that by which he may obtain the greatest amount of necessities,  
4 conveniences, and luxuries, with the smallest quantity of labour and physical self-denial with  
5 which they can be obtained in the existing state of knowledge” (Mill 1967, p.326). Of course,  
6  
7  
8 Mill never considered increasing wealth and economic growth as society’s ultimate end. And  
9  
10 whatever this aim may be, it is not for political economists to uncover it. This is the task of art,  
11  
12 like moral philosophy, not of science. Mill’s economic man has nevertheless been attacked by a  
13  
14 number of thinkers such as Emile Durkheim (Steiner & Gillig 2015). Durkheim was not averse  
15  
16 to the idea of using abstractions for scientific purpose and he certainly belonged to those who  
17  
18 believed in the scientific potential of social studies, but he believed that Mill’s process of  
19  
20 abstraction went too far. If Mill’s political economy had been supposed to stand on its own,  
21  
22 Durkheim’s criticism would probably have been acceptable. However, this was never supposed  
23  
24 to be the case, and once articulated with ethology, Durkheim’s point falls through.  
25  
26  
27  
28  
29

30  
31 The scientific legitimacy of political economy relies, as presented above, on a de-socialized view  
32  
33 of man. The practical validity of the general laws it has uncovered depends, however, on the  
34  
35 social and historical circumstances actually encountered. Circumstances can strongly affect  
36  
37 people’s character in one way or another and this has potentially decisive consequences on their  
38  
39 economic choices. In the absence of a science capable of anticipating the impact caused on  
40  
41 individual character by institution changes, political economy is condemned to remain a purely  
42  
43 theoretical science, devoid of any possible applications. To be fully relevant, the lessons from  
44  
45 political economy need to be completed by those of ethological principles, the omission of which  
46  
47  
48 “is no defect in them as abstract or hypothetical sciences, but it vitiates them in their practical  
49  
50 application as branches of a comprehensive social science.” (Mill [1843] 1974, p.487). General  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 economic laws ought to be reinterpreted according to social and cultural specificities to have  
4  
5 some ‘applications to social philosophy’ as stated in the full title of the *Principles*.  
6  
7

8  
9 As we know, Mill abandoned the idea of writing a treatise on ethology, and dedicated the few  
10  
11 years immediately following the *System of Logic* to the redaction of his book on political  
12  
13 economy instead. This may explain why he did not strictly follow the epistemological rules he  
14  
15 ascertained in the *Definition of Political economy* (1836) and in the *System* (1843). In the  
16  
17 following sections, I shall present two economic issues: land properties and colonies, where Mill  
18  
19 voluntarily stepped aside from using the abstraction of the *homo oeconomicus* to include  
20  
21 considerations drawn from ethology in order to promote moral improvement.  
22  
23  
24  
25  
26  
27

## 28 29 V. LAND PROPERTY AND COLONIES 30

31  
32 Let us first consider the issue raised by large agricultural exploitation. Mill considers small  
33  
34 landed properties to be far superior to large exploitations, yet he recognizes that, economically  
35  
36 speaking, labour “is unquestionably more productive on the system of large industrial  
37  
38 enterprises” (Mill [1848] 1965, p.768). In chapter nine of his *Principles*, ‘Of Production on a  
39  
40 Large, and Production on a Small Scale’, Mill compares the productivity of small and large  
41  
42 agricultural enterprises and endeavours – not without difficulties – to present the economic  
43  
44 benefits of small agricultural exploitations. He eventually admits that “the comparative merits of  
45  
46 the *grande* and the *petite* culture, especially when the small farmer is also the proprietor, cannot  
47  
48 be looked upon as decided. It is a question on which good judges at present differ. The current of  
49  
50 English opinion is in favour of large farms: on the Continent, the weight of authority seems to be  
51  
52 on the other side” (Mill [1848] 1965, p.150). Property is the determinant factor to take into  
53  
54  
55  
56  
57

1  
2  
3 consideration: “To understand the subject, it must be studied where the cultivator is the  
4 proprietor, or at least a *métayer* with a permanent tenure; where the labour he exerts to increase  
5 the produce and value of the land avails wholly, or at least partly, to his own benefit and that of  
6 his descendants.” (Mill [1848] 1965, p.148).  
7  
8  
9  
10  
11

12  
13 Although not as efficient as hired labour, economic independence gives a sense of responsibility  
14 and prudence to the farmers, whose survival now only depends on their ability to work, save and  
15 plan ahead for their future. In addition to being constitutive of strong and active personalities,  
16 economic independence is also a prerequisite to moral improvement, including sexual restraint.  
17 This applies to rich countries, in which increased production ceases to be an important object  
18 (Mill [1848] 1965, p.755), as well as poor countries, who should give priority to character  
19 formation over immediate economic growth: “In a backward state of industrial improvement, as  
20 in Ireland, I should urge its introduction, in preference to an exclusive system of hired labour; as  
21 a more powerful instrument for raising a population from semi-savage listlessness and  
22 recklessness, to persevering industry and prudent calculation.” (Mill [1848] 1965, p.768). In a  
23 series of articles on Ireland, published in the *Morning Chronicle*, Mill also claims that the Irish  
24 peasants’ admittedly revolting behaviour is the result of their exploitation and that with  
25 proprietorship “tendencies in the current character of the peasant are reversible” (Carlisle 2010,  
26 p.149).  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

46 In the case of Ireland, Mill actually conjectured that the Irish character had been ‘malformed’ by  
47 the economic exploitation and religious subservience imposed by Britain (Bell 2010, p.49). To  
48 put it right, Mill proposed to buy up the wastelands of Ireland and distribute them to the Irish  
49 peasantry, either through outright grants-in-deed, or through permanent tenures under a quit-rent  
50 (Zastoupil 1983, p.710). In India, where the government “being itself the landlord” (Mill [1848]  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 1965, p.327), redistributing land to Indian peasants is not an option. He therefore proposes to  
4 reform the terms of tenure and grant perpetual leases to the farmers. But, whatever the means  
5 employed, changing a national character always remains a long and slow process. A more direct  
6 and efficient way is to aim at individual characters. This perfectly illustrates the principle  
7 enunciated above, and reaffirmed in the *Principles*, according to which “the moral aspect of the  
8 question [...] is still more important than the economical” (Mill [1848] 1965, p.768). His  
9 argument is moral rather than economic, and it relies on the types of psychological mechanisms  
10 ethology should have uncovered. His aim is to foster strong and energetic characters. As noted  
11 by Joseph Persky, “the point was not to redistribute, but to reorganize in a fashion that increased  
12 efficiency and overcame rural lethargy” (Persky 2016, p.106).  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26

27 Mill, like his father before him, believed that England, as one of the most morally advanced  
28 countries in the world, had a duty to civilize underprivileged countries. Unlike his father,  
29 however, and in discordance with the Ricardian tradition, Mill argues that England would gain  
30 from carrying on its ancient colonial trade system with India. Colonies are, for Mill, “the last of  
31 the counter-forces which check the downward tendency of profits, in a country whose capital  
32 increases faster than that of its neighbours, and whose profits are therefore nearer to the  
33 minimum.” (Mill [1848] 1965, pp.745-46). His main argument, developed in the *Principles of*  
34 *Political Economy*, is that only an emigration of capital (as well as population) could check the  
35 progressive decline of the rate of profit and keep up a level “a supply of cheap food and cheap  
36 materials of clothing, proportional to the increase of our population; thus enabling an increasing  
37 capital to find employment in the country, without reduction of profit, in producing  
38 manufactured articles with which to pay for this supply of raw produce.” (Mill [1848] 1965,  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 p.746). If one does accept the wage-fund theory, as Mill does<sup>9</sup>, the reasoning is perfectly valid.  
4  
5 However, it should not be read as a personal stance in favour of imperialism.  
6  
7

8 The reason for which Mill still favours colonization is not economic but moral. Political  
9  
10 economy provides some arguments in favour of imperialism but they are not and cannot be  
11  
12 determinant. It is fundamental, to fully understand Mill's approach to political economy, to keep  
13  
14 in mind the way he sees the relation between science and art (Riley 1998, Robson 1968).  
15  
16

17 Amongst the main sciences, Mill gives a special importance to moral sciences such as political  
18  
19 economy, ethology, history, and psychology, which he discusses in book six of the *System of*  
20  
21 *Logic*. By art, he essentially refers to morality, politics and aesthetics, which form the basis of  
22  
23 the 'art of life'. For Mill, art and science both contribute to improve man's condition but each  
24  
25 within its own constitution. The object of science is to uncover the general laws of the organic  
26  
27 and inorganic world while the aim of art is to set human ends. Science is then strictly  
28  
29 instrumental to art: "The art proposes to itself an end to be attained, defines the end, and hands it  
30  
31 over to the science. The science receives it, considers it as a phenomenon or effect to be studied,  
32  
33 and having investigated its causes and conditions, sends it back to art with a theorem of the  
34  
35 combination of circumstances by which it could be produced." (Mill [1843] 1974, pp.944-45).  
36  
37  
38

39 Despite this important methodological principle, Mill also, on occasion, introduces moral and  
40  
41 normative considerations in his *Principles of Political Economy*. This happens, in particular,  
42  
43 when oppositions occur between potential economic benefits and character improvement. Take  
44  
45 the case of productivity and land reform. As a science, the role of political economy is to analyse  
46  
47 the mechanisms behind economic growth, however, such analyses do not suppose that economic  
48  
49  
50  
51  
52  
53

---

54  
55 <sup>9</sup> This point is, however, debated. See (White 1994, Vint 1995).  
56  
57

1  
2  
3 growth is man's ultimate end. In fact, morality tells us that it is not and that when a conflict  
4 occurs between increase of wealth and improvement of character, priority should be given to the  
5 latter. The difficulty arises when political economy is to give practical instruction on how to  
6 achieve ends set by art. This cannot be done without the support of ethology.  
7  
8  
9  
10  
11

12  
13 Colonialism constitutes a perfect illustration of Mill's belief that "the moral aspect of the  
14 question [...] is still more important than the economical" (Mill [1848] 1965, p.768)<sup>10</sup>. Contrary  
15 to what has sometimes been argued, Mill did not use political economy to justify colonialism.  
16 Like his father, he supported the English colonial system but essentially for moral – or what  
17 Sullivan called 'cultural' – reasons. Colonies, when suitably governed, could benefit from the  
18 experience and wisdom of the old countries. They also, as Bell judiciously noted, constituted  
19 "laboratories of character development" or "case studies of his proposed science of ethology"  
20 (Bell 2010, p.46). This remains true whether colonies change indigenous characters for the better  
21 or the worse. And, as shown by Bell, Mill grew more and more disillusioned about England's  
22 ability to foster strong and noble characters abroad.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36

37 In his *Autobiography*, Mill explains that, though he sometimes took the historical and cultural  
38 context of the economic agent, the *Principles of Political Economy* "yielded to none of its  
39 predecessors in aiming at the scientific appreciation of the action of these causes, under the  
40 conditions which they presuppose; but it set the example of not treating those conditions as final"  
41 (Mill [1874] 1981, p.257). Land properties and colonies are but two examples of these  
42  
43  
44  
45  
46  
47  
48  
49

---

50 <sup>10</sup> It can, however, be pointed out that, as Stefan Collini suggests, Mill often took advantage of  
51 his authority as a leading political economist to influence public opinion on the benefits of  
52 reforms that could not be vindicated by economic arguments (Collini 1991, p.163).  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 provisional conditions. It is possible that Mill broke his own epistemological rule because he  
4 knew his treatise on ethology would never be completed. He may also simply have considered  
5 the issue too politically sensitive not to be dealt with<sup>11</sup>. As such they do not jeopardize the  
6 scientific status of political economy as he defined it. However, I argue that by systematically  
7 promoting individual competition, political economy goes beyond its scientific scope, and that it  
8 actually aims at improving people character.  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

## 21 VI. WHEN COMPETITION SHAPES CHARACTER

22  
23  
24 In what follows, and to close this discussion, I want to show that not only did ethology  
25 sometimes bias Mill's economic reasoning, but that it may also have influenced his defence of  
26 competition.  
27  
28  
29  
30

31 As Smith and Ricardo's successor, Mill supports economic competition, then referred as the  
32 *laissez-faire* principle, as the most effective system imagined so far. We have seen that Mill had  
33  
34  
35

---

36  
37  
38 <sup>11</sup> Mill interrupted the redaction of the *Principles on Political economy* for six months to write  
39 papers about the Irish famine: "The Political Economy was far more rapidly executed than the  
40 Logic, or indeed than anything of importance which I had previously written. It was commenced  
41 in the autumn of 1845, and was ready for the press before the end of 1847. In this period of little  
42 more than two years there was an interval of six months during which the work was laid aside,  
43 while I was writing articles in the *Morning Chronicle*... urging the formation of peasant  
44 properties on the waste lands of Ireland. This was during the period of the Famine, the winter of  
45 1846-47." (Mill [1874 1981, p.242)  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 reasons to also support colonialism. This is not, however, the only exceptions he makes to the  
4 liberal creed. In Book five Chapter eleven of the *Principles*, ‘On the grounds and limits of the  
5 *Laissez-faire* or non-interference principle’, Mill presents an impressive list of exceptions to the  
6 rule of competition.  
7  
8  
9

10  
11  
12  
13 It is often believed, maybe due to the influential thesis of Friedrich Hayek, that Mill’s liberalism  
14 wore off under the spell of his wife Harriet Taylor, who eventually led him away from economic  
15 liberalism (Légé 2008). This is partially incorrect. First, historians have shown that Mill’s early  
16 positions were largely shaped by the Saint-Simonian philosophy (Gillig & Légé 2017). His faith  
17 in the utility of public intervention, however, steadily declines after the publication of  
18 Tocqueville’s *Democracy in America* and the 1848 Revolution. It is also wrong to conclude from  
19 his late socialist sympathies (which actually were limited to local experiments) the idea that he  
20 then came to question the legitimacy of competition. On the contrary, competition always had  
21 been a bone of contention between him and them (Mill [1848] 1965, p.792).  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

34  
35 Whilst writing the *Principles of Political Economy*, between 1845 and 1848, Mill is hesitant  
36 about the importance to give to the laissez-faire principle. He confides his doubts in a letter to  
37 John Austin (Schwartz 1972, p.114) before finally deciding to support the rule of free  
38 competition with a few exceptions on account that “if competition has its evils, it prevents  
39 greater evils.” (Mill [1848] 1965, p.793). Mill is also concerned about the consequences of the  
40 development of democracy (according to Tocqueville’s meaning) in English society. Individual  
41 freedom and competition are, alongside an appropriate political constitution, the major means to  
42 stop the rise of uniformity and apathy characteristic of such societies: “The public collectively is  
43 abundantly ready to impose, not only its generally narrow views of its interests, but its abstract  
44 opinions, and even its tastes, as laws binding upon individuals. [...] there was never more  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 necessity for surrounding individual independence of thought, speech, and conduct, with the  
4 most powerful defences, in order to maintain that originality of mind and individuality of  
5 character, which are the only source of any real progress” (Mill [1848] 1965, p.945).  
6  
7

8  
9  
10 In the *Principles of Political Economy*, Mill lists four reasons supporting the system of free  
11 competition. The first one is related to the imperative right that individuals have to be in control  
12 of their own personal life. According to Mill, “there is a circle around every individual human  
13 being which no government, be it that of one, of a few, or of the many, ought to be permitted to  
14 overstep: there is a part of the life of every person who has come to years of discretion, within  
15 which the individuality of that person ought to reign uncontrolled either by any other individual  
16 or by the public collectively” (Mill [1848] 1965, p.943). But, as Mill acknowledges, this  
17 objection seldom applies to governmental interferences which are generally not directed against  
18 individual free agency. The second objection raised by Mill against the increase of  
19 governmental agency follows from the impact such an increase has on the actual power of the  
20 government, both in terms of authority and influence. The danger is especially when members of  
21 the government have social or individual vested interests to pursue. This is also particularly  
22 alarming in democratic societies in which public opinion absolutely rules and minorities can  
23 easily be oppressed. The third objection against governmental agency is linked to the principle of  
24 the division of labor, or more exactly its implicit denial. The more governments take over what  
25 was previously held by individuals or local powers, the less efficient they are in fulfilling their  
26 functions. This is especially true of centralized countries such as France. The fourth and last  
27 objection but also, according to Mill, the ‘strongest’ one, is the loss of individual agency that  
28 goes along with the growing individual apathy produced by sprawling governmental  
29 interferences: “A people among whom there is no habit of spontaneous action for a collective  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 interest – who look habitually to their government to command or prompt them in all matters of  
4  
5 joint concern – who expect to have everything done for them, except what can be made an affair  
6  
7 of mere habit and routine – have their faculties only half developed; their education is defective  
8  
9 in one of its most important branches.” (Mill [1848] 1965, pp.948-49). This main reason for  
10  
11 governments not to interfere in private lives is that they have disastrous effects on the formation  
12  
13 of characters. The less competitive is a society, the less strong and energetic individual  
14  
15 characters will be. Interestingly enough, competition is not supported by strictly economic  
16  
17 arguments.  
18  
19  
20  
21

22 Despite those reasons, Mill concedes that a number of exceptions can be made to the principle of  
23  
24 competition, which are also developed in chapter eleven, book five, of the *Principles of Political*  
25  
26 *Economy*. The consumer is, theoretically, the best judge of his needs and of his interest.

27  
28 However, this is not always the case: “there are other things, of the worth of which the demand  
29  
30 of the market is by no means a test; things of which the utility does not consist in ministering to  
31  
32 inclinations, nor in serving the daily uses of life, and the want of which is least felt where the  
33  
34 need is greatest. This is peculiarly true of those things which are chiefly useful as tending to raise  
35  
36 the character of human beings. The uncultivated cannot be competent judges of cultivation”  
37  
38

39  
40 (Mill [1848] 1965, p.953). In other words, those who may be the most in need of education are  
41  
42 not necessarily the ones who are the most desirous of it. Education, for Mill, can legitimately be  
43  
44 enforced and paid for by the government, as it is essential to the improvement of human  
45  
46 character. There are a number of further cases for which Mill believes the principle of free  
47  
48 competition should not apply, including child and female labor, perpetual commitments, and  
49  
50 unions. Although each could be linked to character improvement, they are nonetheless justified  
51  
52 otherwise. The last case, a defense of public charity, is ethology related.  
53  
54  
55  
56  
57

1  
2  
3 The case of public charity, or the Poor Laws, was particularly sensitive in the first half of the  
4 nineteenth century in England (Clement 1999). Ricardo, in particular, had voiced his opposition  
5 to such laws, which he considered partly responsible for the poverty of the lowest classes. In  
6 addition to deterring people from work, public charity interfered with the labor market and  
7 enabled employers to pay their workers a rate below the natural wage. Despite adopting very  
8 similar principles to Ricardo, John Stuart Mill is one of the very few classical economists to have  
9 given his support to the Poor Laws. His reasons are not, however, strictly economic but  
10 ethologic<sup>12</sup>. The legitimacy of public assistance ought to be assessed by the consequences it has  
11 on individual character. Mill admits that too generous an assistance can foster idleness<sup>13</sup> but he  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

---

26  
27 <sup>12</sup> Gregory Claeys claims that Mill's views on poor relief can also be explained from a political  
28 standpoint. The idea is that beyond the perfection of individual character lies another, and much  
29 more important goal, the perfection of the social entity (Claeys 2013).  
30  
31  
32  
33

34 <sup>13</sup> Mill considered man to be naturally inclined to indolence and reproached the socialists for  
35 being too naive about it. He sees competition as a necessary condition to exertion: "It is the  
36 common error of Socialists to overlook the natural indolence of mankind; their tendency to be  
37 passive, to be the slaves of habit, to persist indefinitely in a course once chosen. Let them once  
38 attain any state of existence which they consider tolerable, and the danger to be apprehended is  
39 that they will thenceforth stagnate; will not exert themselves to improve, and by letting their  
40 faculties rust, will lose even the energy required to preserve them from deterioration.  
41  
42  
43  
44  
45  
46  
47  
48  
49

50 Competition may not be the best conceivable stimulus, but it is at present a necessary one, and  
51 no one can foresee the time when it will not be indispensable to progress." (Mill [1848] 1965,  
52 p.793)  
53  
54  
55  
56  
57

1  
2  
3 also believes that not being assisted can also have detrimental effects on character: “Energy and  
4 self-dependence are (...) liable to be impaired by the absence of help, as well as by its excess. It  
5  
6 is even more fatal to exertion to have no hope of succeeding by it, than to be assured of  
7  
8 succeeding without it. When the condition of any one is so disastrous that his energies are  
9  
10 paralyzed by discouragement, assistance is a tonic, not a sedative: it braces instead of deadening  
11  
12 the active faculties” (Mill [1848] 1965, pp.967-68). The difficulty is to find a system that  
13  
14 supports people in real need and prevent them from giving up whilst not discouraging those who  
15  
16 do not need support but seek it. This implies differentiating assistance according to the situation  
17  
18 and character of the people calling for help. Private individuals only occasionally meet people in  
19  
20 need and they tend to act according to their emotions rather than their reason. They either give  
21  
22 too much or too little. When properly managed, public charity, on the contrary, can be efficient;  
23  
24 and Mill believes history has proven English Poor Laws to be effective (Mill [1848] 1965,  
25  
26 p.968).  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36

## 37 VII. MILL ON PATERNALISM

38  
39  
40 We have seen that education was, for Mill, one of the most efficient ways to mould characters.  
41  
42 But, providing that it equally applies to all, it is also one of the slowest. “The initiation of all  
43  
44 wise or noble things”, claims Mill, “comes and must come from individuals” (Mill [1859] 1977,  
45  
46 p.269). Being altruistic, noble characters are logically incited to assist the less privileged to rise  
47  
48 to higher moral and hedonistic standards. But more than being a mere incentive to altruistic  
49  
50 people, inequality of characters, caused as previously seen by adverse social circumstances, are  
51  
52 also a matter of social justice.  
53  
54  
55  
56  
57

1  
2  
3 In the absence of external guidance, people born in unfavourable circumstances cannot hope to  
4 reach high levels of happiness and morality. It is therefore essential for Mill that those who were  
5  
6 lucky enough to benefit from better circumstances therefore help those less fortunate to  
7  
8 overcome their condition. And as it happens, they are the only ones to be in a position to assist  
9  
10 them. Noble characters have benefited from privileged instruction and education. They now have  
11  
12 a moral duty towards those who have not: “Human beings owe to each other help to distinguish  
13  
14 the better from the worse, and encouragement to choose the former and avoid the latter. They  
15  
16 should be for ever stimulating each other to increased exercise of their higher faculties, and  
17  
18 increased direction of their feelings and aims towards wise instead of foolish, elevating instead  
19  
20 of degrading, objects and contemplations.” (Mill [1859] 1977, p.277).  
21  
22  
23  
24  
25  
26  
27

28 For Mill, there is nothing wrong in accepting 'guidance' from more privileged people: “Many  
29  
30 have let themselves be guided (which in their best times they always have done) by the counsels  
31  
32 and influence of a more highly gifted and instructed One or Few.[...] The honour and glory of  
33  
34 the average man is that he is capable of following that initiative; that he can respond internally to  
35  
36 wise and noble things, and be led to them with his eyes open.” (Mill [1859] 1977, p.269).  
37  
38  
39

40 Individuals are, for Mill, all progressive beings and as such they have a decisive part to play in  
41  
42 their own moral development. Although it is always desirable for an individual to reform his own  
43  
44 character, it is in practice hardly possible for him to do so in isolation. Reformation of character  
45  
46 can only be indirect: “I can indeed influence my own volitions, but only as other people can  
47  
48 influence my volitions, by the employment of appropriate means. Direct power over my volitions  
49  
50 I am conscious of none.” (Mill [1865] 1979, p.298-99). In other words, reformation of one's  
51  
52 character cannot be – strictly speaking – a self-regarding action.  
53  
54  
55  
56  
57

1  
2  
3 Mill reckons, however, that some individuals are better equipped than others to initiate these  
4 actions. “It is no novelty”, he writes in *Subjection of Women*, “that mankind do not distinctly  
5 foresee their own changes, and that their sentiments are adapted to past, not to coming ages. To  
6 see the futurity of the species has always been the privilege of the intellectual *élite*, or of those  
7 who have learnt from them; to have the feelings of that futurity has been the distinction, and  
8 usually the martyrdom, of a still rarer *élite*.” (Mill [1869] 1984, 294). They belong to what  
9 Samuel Taylor Coleridge originally called *clerisy*, a group of men of superior intellectual and  
10 moral standards. It is likely, therefore, that this particular feature of Mill’s philosophy is a  
11 testament to the Romantics’ deep influence exercised during his youth. But, as argued by  
12 Capaldi, the nature of Coleridge’s notion of *clerisy* changed with Mill, most probably as a result  
13 of the influence of Comte (Capaldi 2004, p.170). For Maurice Cowling, the *clerisy* performs the  
14 functions once performed by the clergy, substitution for the ideal object of Christianity which is  
15 altruistic elevation of feelings (Cowling 1963, p.78). It is unclear, however, what Mill really  
16 expects from this group and how they could be identified. Could the members of the *clerisy* be  
17 those granted with multiple votes in *Representative Government*? To what extent are these  
18 individuals entitled to interfere with other people’s lives? With what legitimacy? Mill explicitly  
19 rejects any form of paternalism. Legal paternalism is explicitly discarded by his famous harm  
20 principle. Individuals, however, can, under some conditions, interfere with other people’s lives.  
21 “It would be a great misunderstanding of this doctrine”, explains Mill in *On Liberty*, “to suppose  
22 that it is one of selfish indifference, which pretends that human beings have no business with  
23 each other's conduct in life, and that they should not concern themselves about the well-doing or  
24 well-being of one another, unless their own interest is involved. Instead of any diminution, there  
25 is need of a great increase of disinterest exertion to promote the good of others. But disinterested

1  
2  
3 benevolence can find other instruments to persuade people to their good, than whips and  
4  
5 scourges, either of the literal or metaphorical sort.” (Mill [1859] 1977, p.277).  
6

7  
8 The first of these conditions is to never use coercion or threat, but rather to use persuasion, as  
9  
10 described in the last, but rarely quoted, sentence of the harm principle: “That the only purpose  
11  
12 for which power can be rightfully exercised over any member of a civilized community, against  
13  
14 his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient  
15  
16 warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to  
17  
18 do so, because it will make him happier, because, in the opinions of others, to do so would be  
19  
20 wise, or even right. These are good reasons for remonstrating with him, or persuading him, or  
21  
22 entreating him, but not for compelling him, or visiting him with any evil in case he do  
23  
24 otherwise.” (Mill [1859] 1977, p.224). This passage, considered *in extenso*, demonstrates the  
25  
26 misconception that subsequently arose from an over-interpretation of the expression ‘against his  
27  
28 will’.  
29  
30  
31

32  
33 For private actions on character formation to take place, secondly, they ought not to express or  
34  
35 reinforce any existing subjection relationship between individuals. This notably applies to  
36  
37 relations between rich and poor, between different social classes, between men and women, or  
38  
39 between different races. This point is defended in chapter seven, book four, of the *Principles of*  
40  
41 *Political Economy*. Mill is appalled by the idea that a category of individuals, essentially  
42  
43 composed of rich men, sees itself as being *in loco parentis* to the poor and women, “guiding and  
44  
45 restraining them like children” (Mill [1848] 1965, p.759). It relies on a strongly idealized  
46  
47 concept of the relationship between protector and protected, which conceals its real authoritative  
48  
49 nature. “The so-called protectors are now the only persons against whom, in any ordinary  
50  
51 circumstances, protection is needed”, contends Mill (Mill [1848] 1965, p.761).  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 This leads me, to finish, to question a possible similarity between Mill's project of ethology and  
4 Cass Sunstein and Richard Thaler's libertarian paternalism. Although the authors reject any  
5 intellectual legacy from the English philosopher, it is impossible not to notice a certain similarity  
6 with the way they respectively propose to use psychology and behavioural economics to improve  
7 people's lives. There is a striking similarity between Mill's ethology and Sunstein and Thaler's  
8 use of empirical studies in behavioural economics. Considering Mill's emblematic figure in the  
9 debates over paternalism, I believe the question of its potential connection with libertarian  
10 paternalism both offers a valuable criticism of this new trend in paternalism as well as a better  
11 idea of Mill's originality.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

24 Over the last twenty years a number of books and articles have been published in defence of  
25 what Sunstein and Thaler coined libertarian paternalism (Sunstein and Thaler 2003), that is to  
26 say a form of paternalism which is fully compatible with the principles of individual freedom.  
27 Based on a vast literature of experimental psychology studies, libertarian paternalism, or  
28 behavioural paternalism as it is also referred (Salvat 2014, Lecouteux 2016), argues that since  
29 individual choices are unavoidably subject to other people's choices, the latter actors are  
30 therefore obliged to take the welfare of the former into account. This can be achieved, for  
31 instance, through carefully designed default rules, framing or endowment effects (Loewenstein  
32 and Haisley 2008). Apart from its literary and scientific success, behavioural paternalism has  
33 been experimented in traditionally liberal countries such as the USA and the United Kingdom,  
34 where nudges units are or have been implemented seen.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

49 For Sunstein and Thaler, all paternalistic actions are legitimate as long as they promote the  
50 'right' choice and that individuals can freely turn them down. In practice, they believe  
51 paternalistic actions are needed in all kinds of decisions in people's lives, from pension schemes  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 to grocery shopping, considering that “almost all people, almost all the time, make choices that  
4 are [not] in their best interest” (Sunstein and Thaler 2003, 1163). More than being merely  
5  
6 justifiable, they claim that behavioural paternalism is in fact unavoidable. Since individuals are  
7  
8 still “free to choose”, the authors argue that their version of paternalism is fully compatible with  
9  
10 libertarianism. Like Mill, Sunstein and Thaler consider that there is an objective good but they  
11  
12 plainly disagree about its nature. For Mill, the objective good translates into being of noble  
13  
14 character. It covers both a sensibility to higher pleasures and some degree of selflessness. For  
15  
16 Sunstein and Thaler, on the contrary, the goodness of an action exclusively relates to the  
17  
18 objective welfare of the agent.  
19  
20  
21  
22  
23  
24  
25  
26

## 27 VIII. CONCLUSION

28  
29  
30 I have presented in this paper Mill’ projected science of ethology, the aim of which was to  
31  
32 uncover the psychological laws of individual behavior. Mill believed that such laws could then  
33  
34 be used to improve people’s character. Although he never finished the treaty of ethology as he  
35  
36 had planned, I argue that this project resurfaces in his work on political economy, as well as in  
37  
38 some of his writings on politics and justice. The fact that Mill was indeed ready to interfere with  
39  
40 people’s lives for their own good is fundamental and invites a reassessment of his views on  
41  
42 paternalism. It also gives us invaluable tools to reflect on the limits of a contemporary form of  
43  
44 liberal yet psychology-based form paternalism, namely libertarian paternalism.  
45  
46  
47

48 Mill’s paternalistic views constitute, what I believe to be, the best argument against modern  
49  
50 versions of paternalism, namely their lack of truly moral ends. Behavioral paternalism justifies  
51  
52 paternalistic actions on account of the welfare improvement it will bring to individuals but this is  
53  
54 not a sufficient reason for Mill to interfere with private lives. The authors neglect what  
55  
56  
57

1  
2  
3 constitutes for Mill the moral core of his paternalism, that is man's betterment. Mill considers  
4 that the only real objective of paternalism is to improve man's moral character, rather than the  
5 immediate outcome of his actions. Contrary to Sunstein and Thaler, Mill is concerned by man's  
6 ability to choose for himself and he would certainly regard nudges as a further debasement of  
7 man's character and individuality. For Mill, education is – along with the use of colonialism, in  
8 some instances – the most effective way to improve people's character. Additionally, private and  
9 benevolent actions are possible in some scenarios. The only condition expressed by Mill is that  
10 people's autonomy and diversity of character is promoted. The existence of an objectively  
11 superior good goes with the idea that there are many ways to achieve it and that it cannot be  
12 simply expressed by an economic aggregate. If paternalism is justified at all, it should be moral  
13 in its core.  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

## 32 REFERENCES

- 33  
34  
35 Albee, Ernest. 1902. *A History of English Utilitarianism*, London: Thoemmes Press.  
36  
37 Arneson, Richard. 1980. "Mill versus Paternalism." *Ethics* 90 (4): 470-489.  
38  
39 Bain, Alexander. 1882. *James Mill: A biography*. Longmans: Green, and co.  
40  
41 Ball, Terence. 2000. "The Formation of Character: Mill's "Ethology" Reconsidered.", *Polity* 33  
42 (1): 25-48.  
43  
44  
45  
46 Bell, Duncan. 2010. "John Stuart Mill on Colonies." *Political Theory* 38 (1): 34-64.  
47  
48 Capaldi, Nicholas. 2004. *John Stuart Mill: A Biography*. Cambridge: Cambridge University  
49 Press.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 Capaldi, Nicholas. 1973. "Mill's Forgotten Science of Ethology." *Social Theory and Practice* 2  
4  
5 (4): 409-420.  
6  
7  
8 Carlisle, Janice. 2010. *John Stuart Mill and the Writing of Character*. Athens: University of  
9  
10 Georgia Press.  
11  
12 Claeys, Gregory. 2013. *Mill and Paternalism*, Cambridge: Cambridge University Press.  
13  
14 Clément, Alain. 1999. *Nourrir le peuple, entre État et marché, XVI siècle- XIX siècle*, Paris:  
15  
16 L'Harmattan.  
17  
18  
19 Collini, Stefan. 1991. *Public moralists: political thought and intellectual life in Britain, 1850-*  
20  
21 *1930*. Oxford: Clarendon Press.  
22  
23  
24 Comte, Auguste. 1988. *Introduction to Positive Philosophy*, Translated by F. Ferré,  
25  
26 Indianapolis/Cambridge: Hackett Publishing Company.  
27  
28  
29 Constant, Benjamin. 1988. *Political Writings*. Cambridge: Cambridge University Press.  
30  
31 Cowling, Maurice. 1963. *Mill and Liberalism*. Cambridge: Cambridge University Press.  
32  
33 Dworkin, Gerald. 1997. *Mill's On Liberty: Critical Essays*, Rowman & Littlefield Publishers.  
34  
35 Fuchs, Alan E. 2001. "Autonomy, Slavery, and Mill's Critique of Paternalism." *Ethical Theory*  
36  
37 *and Moral Practice*, 4(3): 231-251.  
38  
39  
40 Gillig, Phippe & Légé, Philippe. 2017. "De la défiance à l'éloge des coopératives par J. S. Mill :  
41  
42 retour sur la constitution d'une pensée libérale dans la première moitié du XIXe siècle."  
43  
44 *Cahiers d'économie Politique / Papers in Political Economy*, 73(2): 197-221.  
45  
46  
47 Gould, Steven. 1996. *The Mismeasure of Man*, New York: Norton.  
48  
49  
50 Harrison, John. 2009. *Robert Owen and the Owenites in Britain and America (Routledge*  
51  
52 *Revivals): The Quest for the New Moral World*. London: Taylor & Francis.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 Hollander, Samuel. 1985. *The Economics of John Stuart Mill: Political economy*. Vol. 2,  
4  
5 Toronto, University of Toronto Press.  
6  
7  
8 Houghton, Walter E. 1957. *The Victorian Frame of Mind, 1830-1870*. New Haven and London:  
9  
10 Yale University Press.  
11  
12 Leary, David E. 1982. "The Fate and Influence of John Stuart Mill's Proposed Science of  
13  
14 Ethology." *Journal of the History of Ideas*, 43: 153-162.  
15  
16  
17 Lecouteux, Guilhem. 2016. "From Homo Economicus to Homo Psychologicus: the Paretian  
18  
19 Foundations of Behavioural Paternalism." *Æconomia* 6 (2): 175-200.  
20  
21  
22 L  g  , Philippe. 2008. "Hayek's Readings of Mill." *Journal of the History of Economic Thought*.  
23  
24 Volume 30 (2): 199-215.  
25  
26  
27 Loewenstein, George .F., and Emily Haisley. 2008. "The Economist as Therapist:  
28  
29 Methodological Ramifications of 'Light' Paternalism." In Andrew Caplin and Andrew  
30  
31 Schotter, eds., *The Foundations of Positive and Normative Economics. A Handbook*, New  
32  
33 York: Oxford University Press, pp.210-245.  
34  
35  
36 Mill, John Stuart. [1848] 1965. *Principles of Political Economy with some of their Applications*  
37  
38 *to Social Philosophy, The Collected Works of John Stuart Mill*, ed. John M. Robson,  
39  
40 Volumes II & III, Toronto: University of Toronto Press.  
41  
42  
43 Mill, John Stuart. [1859] 1967. *On Liberty*. in John M. Robson ed., *Essays on Economics and*  
44  
45 *Society, The Collected Works of John Stuart Mill*, Volume XVIII, Toronto: University of  
46  
47 Toronto Press, pp.213-310.  
48  
49  
50 Mill, John Stuart. 1967. *Essays on Economics and Society, The Collected Works of John Stuart*  
51  
52 *Mill*, Volume XVIII-XIX, ed. John M. Robson, Toronto: University of Toronto Press.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 Mill, John Stuart. [1843] 1974. *A System of Logic Ratiocinative and Inductive. The Collected*  
4  
5 *Works of John Stuart Mill*, Volume VII-VIII, ed. John M. Robson, Toronto: University of  
6  
7 Toronto Press.  
8  
9  
10 Mill, John Stuart. 1977. *Essays on Politics and Society, The Collected Works of John Stuart Mill.*  
11  
12 Volume XVIII-XIX, ed. John M. Robson, Toronto: University of Toronto Press.  
13  
14  
15 Mill, John Stuart. [1865] 1979. *An Examination of William Hamilton's Philosophy and of the*  
16  
17 *Principal Philosophical Questions Discussed in his Writings, The Collected Works of*  
18  
19 *John Stuart Mill.* Volume IX. ed. John M. Robson, Toronto: University of Toronto Press.  
20  
21  
22 Mill, John Stuart. [1874] 1981. *Autobiography and Literary Essay. The Collected Works of John*  
23  
24 *Stuart Mill.* Volume I, ed. John M. Robson, Toronto: University of Toronto Press.  
25  
26  
27 Mill, John Stuart. [1869] 1984. *On Subjection of Women*, in *Essays on Equality, Law, and*  
28  
29 *Education. The Collected Works of John Stuart Mill.* Vol. XXI, by John M. Robson,  
30  
31 Toronto : University of Toronto Press, pp.259-340.  
32  
33  
34 Mill, John Stuart. [1838] 1985. *Bentham*, in John M. Robson, ed., *Essays on Ethics, Religion and*  
35  
36 *Society, The Collected Works of John Stuart Mill.* Volume X, Toronto: University of  
37  
38 Toronto Press, pp.75-115.  
39  
40  
41 Mill, John Stuart. [1861] 1985. *Utilitarianism*, in *Essays on Ethics, Religion and Society, The*  
42  
43 *Collected Works of John Stuart Mill.* Volume X, ed. John M. Robson, Toronto:  
44  
45 University of Toronto Press, pp.203-259.  
46  
47  
48 Olivier, Robert. 2002. "La tentation comtienne de John Stuart Mill : un " disciple indiscipliné."  
49  
50 *Revue d'Histoire des Sciences Humaines* 7 (2), p. 129-156.  
51  
52  
53 Owen, Robert. 1813. *A New View of Society: or, Essays on the Principle of the Formation of the*  
54  
55 *Human Character.* London.  
56  
57  
58  
59  
60

- 1  
2  
3 Riley, Jonathan. 1988. *Liberal Utilitarianism: Social Choice Theory and J. S. Mill's Philosophy*.  
4  
5 Cambridge: Cambridge University Press.  
6  
7  
8 Riley, Jonathan. 1998. "Mill's Political Economy: Ricardian science and liberal utilitarian art." In  
9  
10 John Skorupski, ed., *The Cambridge Companion to Mill*. Cambridge: Cambridge  
11  
12 University Press, pp. 293-337.  
13  
14  
15 Robson, John M. 1998. "Civilization and culture as moral concepts." In John Skorupski, ed., *The*  
16  
17 *Cambridge Companion to Mill*. Cambridge: Cambridge University Press, pp. 338-371.  
18  
19  
20 Ryan, Alan. 2012. *The Making of Modern Liberalism*. Princeton: Princeton University Press.  
21  
22 Ryan, Alan. 1998. "Mill in a liberal lanscape." In John Skorupski, ed., *The Cambridge*  
23  
24 *Companion to Mill*, Cambridge, Cambridge University Press, pp. 497-540.  
25  
26  
27 Salvat, Christophe. 2014. "Behavioral Paternalism." *Revue de Philosophie Economique/ Review*  
28  
29 *of Economic Philosophy* 15 (2): 109-130.  
30  
31 Scharff, Robert C. 1989. "Mill's misreading of Comte on 'interior observation'", *The Journal of*  
32  
33 *the History of Philosophy* 27 (4): 559-572.  
34  
35  
36 Schwartz, Pedro L. 1972. *The New Political Economy of J.S. Mill*. Durham, North Carolina: Duke  
37  
38 University Press.  
39  
40 Skorupski, John. 1989. *John Stuart Mill*. London: Routledge & Kegan Paul.  
41  
42  
43 Sunstein, Cass R., and Richard H. Thaler. 2003. "Libertarian Paternalism Is Not an Oxymoron",  
44  
45 *The University of Chicago Law Review* 70 (4): 1159-1202.  
46  
47  
48 Tait, Gordon. 1971. "Robert Owen on Character Formation and Education.", *Religious*  
49  
50 *Education*, 66(6): 450-456.  
51  
52 Vint, John. 1995. "John Stuart Mill's Wages Fund Recantation: A Lakatosian Analysis." *Journal*  
53  
54 *of Interdisciplinary Economics*, 6(4): 233-254.  
55  
56  
57

1  
2  
3 Wilson, Fred. 1991. "Mill and Comte on the Method of Introspection." *Journal of the History of*  
4 *the Behavioral Sciences* 27 (2): 107-129.  
5  
6

7  
8 White, Michael V. 1994. "'That God-Forgotten Thornton': Exorcising Higglings after On  
9 Labour" in Neil De Marchi and Morgan, Mary S., eds., *Higglings: Transactors and their*  
10 *Markets in the History of Economics, History of Political Economy*, annual supplement,  
11 Durham: Duke University Press, pp.149-83  
12  
13  
14  
15

16  
17 Zastoupil, Lynn. 1983. "Moral Government: J. S. Mill on Ireland." *The Historical Journal* 26  
18 (3): 707-717.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

## Parfit et les raisons du conséquentialisme de la règle

RÉSUMÉ. — *L'article s'interroge sur les relations ambivalentes entre la pensée de Derek Parfit et l'utilitarisme. Dans On What Matters, Parfit rejette l'idée d'une incompatibilité des philosophies morales, arguant que les théories utilitariste, kantienne et contractualiste peuvent être reformulées dans une seule et même « triple théorie ». Cette convergence est-elle réelle ? L'inspiration kantienne et le recours au contractualisme de Tim Scanlon ne laissent-ils pas à l'utilitarisme la portion congrue ?*

ABSTRACT. — *The paper examines the ambivalent relationships between Derek Parfit's thought and utilitarianism. In On What Matters, Parfit rejects the idea of the incompatibility of moral philosophies, arguing that utilitarian, Kantian and contractualist theories can be reformulated into one and the same "triple theory". Is this convergence effective ? What space do this Kantian inspiration and the use of Tim Scanlon's contractualism leave for utilitarianism ?*

---

Parfit a toujours entretenu avec l'utilitarisme une relation assez ambivalente. Son premier ouvrage, *Reasons and Persons*, publié en 1984, a été remarqué à la fois pour ses critiques acerbes du raisonnement utilitariste (notamment *via* la « conclusion répugnante<sup>1</sup> ») et pour sa théorie du réductionnisme ontologique, qui constitue la meilleure réponse apportée à ce jour à l'objection de la séparabilité des personnes, présentée par Rawls dans les années 1970<sup>2</sup>. *On What Matters* permet de mieux situer sa pensée. À première vue, Parfit réfute l'idée même de philosophies morales rivales, arguant que – bien comprises – les théories utilitariste, kantienne et contractualiste peuvent être reformulées dans une seule et même théorie. De fait, Parfit développe une théorie conséquentialiste et hédoniste très influencée par le « fondamentalisme des raisons » du contractualiste Tim Scanlon et par une interprétation kantienne du conséquentialisme de la règle. Mais la convergence des théories est contestable et seulement apparente. Les kantien rejettent la lecture conséquentialiste de Parfit et

1. Voir l'article de Stéphane Zuber dans la présente livraison.

2. Voir C. SALVAT, « Parfit, l'égoïsme rationnel et la question de l'identité personnelle », *Oeconomia*, 5-4 (2015), pp. 437-460.



Scanlon lui-même refuse que sa philosophie soit assimilée à celle de Kant. Quant au conséquentialisme de la règle auquel Parfit parvient, il ne s'agit plus, comme son nom l'indique, d'une forme d'utilitarisme, mais d'une forme de conséquentialisme. Or le conséquentialisme, à la différence de l'utilitarisme, n'est pas nécessairement une philosophie morale (le conséquentialisme n'est éthique que dans la mesure où le critère d'évaluation des conséquences de l'action est un critère moral). Le contractualisme (scanlonien) est compatible avec le conséquentialisme et Parfit affirme que le kantisme est compatible avec le conséquentialisme de la règle. Ni l'une ni l'autre ne peuvent, cependant, converger vers l'utilitarisme. Alors comment expliquer l'insistance de Parfit à défendre la convergence des théories éthiques ? Pourquoi, surtout, les réduire *in fine* à un simple conséquentialisme de la règle lorsque, de toute évidence, l'essentiel de sa construction théorique repose sur les principes du contractualisme ?

*La nature objective des raisons.*

Les raisons, nous dit Parfit, justifient les normes. Il existe traditionnellement deux grandes approches selon que l'on considère ou non le sujet comme la justification ultime de la norme. Selon la première approche, nos raisons sont relatives aux faits satisfaisant (ou appelés à satisfaire) nos désirs. Dans la mesure où l'importance de ces faits n'est relative qu'à nos propres désirs, et donc à notre propre personne, on peut selon Parfit également les appeler raisons relatives au sujet (*subject-given reasons*). Leurs forces dépendent de la force de nos désirs. Selon la seconde approche, les raisons sont relatives aux faits qui concernent les objets de nos désirs. On parle alors de raisons relatives à l'objet (*object-given reasons*). Leurs forces dépendent des faits qui nous donnent des raisons de les désirer. Ajoutons à ces deux catégories un troisième type de raisons relatives à l'état (*state-given reasons*), qui à la différence des raisons relatives à l'objet ne sont pas relatives aux faits qui concernent les objets de nos désirs mais aux états constitués par le fait d'avoir ces désirs<sup>3</sup>. Parfit, à la différence de Scanlon notamment<sup>4</sup>, considère cette dernière catégorie comme secondaire<sup>5</sup>.

3. D. PARFIT, *On What Matters I*, Oxford, Oxford University Press, 2011 (noté désormais *OWM I*), p. 50.

4. Scanlon considère, en effet, que l'évaluation des biens substantiels (il rejette la terminologie de bien objectif) repose, en grande partie, sur la valeur des états de conscience désirables (*desirable states of consciousness*) (T. SCANLON, « Value, desire and Quality of Life », in A. SEN & M. NUSSBAUM, *Quality of Life*, Oxford, Oxford University Press, 1991, pp. 185-200).

5. D. PARFIT, *OWM I*, p. 51.

Quelques précisions sont nécessaires ici. Affirmer qu'une théorie des raisons est objective ne signifie pas, tout d'abord, qu'il s'agisse d'une théorie des raisons objectives. En effet, l'objectivité des raisons ne doit pas être comprise comme renvoyant à un jugement extérieur, un regard objectif, porté sur les raisons de l'individu ; elle consiste simplement en l'extériorité des raisons par rapport au sujet. Parfit confond parfois ces deux perspectives. À strictement parler, les raisons relatives à l'objet ne sont pas nécessairement objectives dans le sens où elles feraient l'objet d'une évaluation totalement décontextualisée. Les raisons relatives à l'objet ne sont pas *subjectives* parce qu'elles sont factuelles, c'est-à-dire qu'elles ont trait aux propriétés de l'objet du désir plutôt qu'au désir lui-même, mais elles n'en sont pas moins *personnelles*, c'est-à-dire que leur valeur dépend du rapport entre l'objet et la personne ainsi que les circonstances dans lesquelles l'évaluation de la force de la raison a lieu. Considérons le fait de manger une pomme, par exemple. Selon la théorie des raisons relatives au sujet, j'ai une raison de manger une pomme si je désire manger une pomme. Les caractéristiques de l'objet (ici la pomme) n'entrent pas en ligne de compte dans la justification de mon action « manger la pomme ». Je pourrais remplacer ici « pomme » par « terre » et justifier mon action de manger de la terre par le fait que je désire manger de la terre. Parfit rejette cette approche. Il considère, au contraire, que ce sont les qualités intrinsèques de la pomme, autrement dit ses qualités nutritives et gustatives qui justifient l'action « manger une pomme ». Mais, bien évidemment, cette raison ne s'applique pas également à tous et dans toutes circonstances. Certaines personnes n'apprécient pas ses qualités gustatives, n'ont tout simplement pas faim (auquel cas la qualité nutritive de la pomme est sans objet) ou encore que la pomme, dans certaines circonstances, n'est pas comestible ou digeste. Dans ces cas-là, comme dans d'autres, il n'est pas justifié pour la personne en question de manger cette pomme (même s'ils désirent la manger).

Ce point est, il me semble, important car il permet de mieux comprendre la nature exacte de l'hédonisme parfitien. Parfit, rappelons-le, distingue trois grandes approches de la valeur : l'hédonisme (selon laquelle ce qui est bon est ce qui donne du plaisir aux gens), la théorie de la satisfaction des désirs (selon laquelle ce qui est bon est ce qui satisfait les désirs d'une personne tout au long de sa vie), et la théorie de la liste objective (selon laquelle certaines choses sont bonnes que nous les désirions ou non)<sup>6</sup>. Dans *On What Matters*, Parfit – on l'a dit – rejette explicitement toute approche subjective des raisons<sup>7</sup>. Il rejette

6. D. PARFIT, *Reasons and Persons*, Oxford, Oxford University Press, 1984, p. 4 (noté désormais *RP*).

7. D. PARFIT, *OWM I*, p. 65.

donc ce qu'il appelle les raisons relatives aux désirs (*desire-based reasons*) au profit des raisons relatives aux valeurs (*value-based reasons*). Mais, et c'est un point essentiel, le rejet des raisons relatives au désir n'implique pas le rejet de l'hédonisme dans la mesure où les raisons hédonistes ne sont pas relatives aux désirs que l'on a mais aux états de plaisirs produits par l'objet du désir<sup>8</sup>. Elles appartiennent donc bien aux raisons relatives à l'objet. L'argument dit de l'agonie (*the agony argument*) pourrait faire croire que Parfit embrasse une théorie de la valeur objective. Ce n'est pas le cas. L'argument de l'agonie, selon lequel nous avons tous des raisons de vouloir éviter une agonie future, vise à décrédibiliser les raisons relatives aux désirs (ce que Parfit appelle également le *subjectivism*), et non pas l'hédonisme. Il est ainsi tout à fait possible d'associer des raisons relatives aux valeurs (ce que Parfit appelle l'*objectivism*) avec des raisons hédonistes relatives à l'objet. Dans une telle perspective, même si on ne se soucie pas de notre futur bien-être, on a des raisons de s'en soucier<sup>9</sup>. Le fait que la souffrance soit mauvaise en tant qu'elle implique des raisons de l'éviter ne signifie pas que la souffrance soit objectivement mauvaise, c'est-à-dire qu'elle puisse être jugée comme telle par une autorité extérieure<sup>10</sup>. La dimension objective de l'hédonisme de Parfit fait donc ici uniquement référence à la qualité des objets du désir de l'individu et non à l'extériorité du jugement. C'est la raison pour laquelle son rejet du subjectivisme n'implique pas son adhésion à la théorie de la liste objective.

Ceci nous permet également de mieux comprendre la distinction établie par Parfit entre le caractère *impersonnel* et le caractère *impartial* des raisons. L'impartialité, en philosophie, est une qualité du jugement moral généralement associée à un point de vue objectif, c'est-à-dire extérieur à l'individu, ce que l'on a appelé le « point de vue de l'univers » (Sidgwick) ou le « point de vue de nulle part » (Nagel). Du fait de cette association, on oppose généralement ce qui est personnel à ce qui est impartial. Pour Parfit, cependant, l'objectivité renvoie, ainsi qu'on l'a dit, à l'objet du désir et non au point de vue adopté. Le caractère personnel d'une raison, c'est-à-dire le fait qu'elle est considérée comme raison d'agir par une personne, n'est par conséquent pas un élément déterminant pour juger de son impartialité, c'est-à-dire de sa conformité à la morale. Il est donc bon de distinguer les raisons *personnelles* (c'est-à-dire les raisons qui s'imposent aux personnes) des raisons *relatives aux personnes* (c'est-à-dire des raisons ayant spécifiquement trait au bien-être de la personne

8. *OWM I*, p. 67.

9. *OWM I*, p. 74.

10. *OWM I*, pp. 371-372.

en question ou à celui de personnes qui lui sont chères)<sup>11</sup>. Pour le dire plus simplement, les raisons personnelles ne sont pas nécessairement intéressées alors que les raisons partiales le sont.

L'idée selon laquelle les raisons peuvent à la fois être personnelles et impartiales est centrale dans l'objectif qui est le sien de résoudre le dilemme du dualisme de la raison pratique de Sidgwick. Ce dernier, rappelons-le, considérerait que nous avons à la fois des raisons d'agir pour le mieux, pensé de façon impartiale, et des raisons d'agir pour ce que nous considérons être notre propre bien, mais que les unes et les autres ne sont pas comparables. Nous avons, en effet, des raisons suffisantes d'agir dans un sens comme dans l'autre, mais aucune raison décisive d'agir dans un sens plutôt que dans un autre. Les deux alternatives, qui expriment respectivement le point de vue objectif (ou, selon l'expression de Sidgwick, « le point de vue de l'Univers ») et le point de vue subjectif, sont toutes deux parfaitement rationnelles.

Pour Parfit, au contraire, les deux points de vue sont concomitants. Le fait que j'adopte un point de vue personnel n'exclut pas le fait que je puisse penser de façon impartiale. Adopter un *point de vue* impartial n'implique pas que l'on soit impartial<sup>12</sup>. Le point de vue personnel n'est en effet pas réductible à l'égoïsme rationnel. Je suis *à la fois* capable de désirer être sauvé et de désirer que le maximum de personnes soient sauvées. Le fait que ces raisons diffèrent, voire entrent en conflit, ne permet pas de conclure, à l'instar de Sidgwick, à l'indétermination du choix. La principale difficulté, pour Parfit, consiste cependant à établir les termes de la comparaison entre les raisons qui sont causées par l'objet d'un désir personnel et celles qui ne le sont pas.

#### *Raisons déontiques vs raisons non déontiques.*

La théorie des raisons adoptée par Parfit doit, on l'a déjà souligné, beaucoup à Scanlon. Or, le philosophe de Harvard conteste le fait que le bien-être individuel puisse être présenté comme le facteur unique, voire même primordial, auquel se réduirait *in fine* toutes les raisons génériques. Une telle approche des raisons conduirait à développer ce qu'il nomme un contractualisme du bien-être (*welfarist contractualism*) auquel il refuse explicitement d'être associé.

11. *OWM I*, p. 40.

12. *OWM I*, p. 41. Inversement, la subjectivité n'est pas incompatible, pour Parfit, avec l'impartialité des raisons. Je peux reconnaître, en tant que sujet, que ma vie ne vaut pas mieux que celle d'un autre (*OWM I*, p. 139).

Parmi les raisons que l'hédonisme ne permet pas d'appréhender sont naturellement « les raisons ayant clairement un caractère moral<sup>13</sup> » dont celles ayant trait à la responsabilité individuelle et à la justice (*fairness*). Parfit institutionnalise en quelque sorte cette distinction en distinguant des raisons déontiques, dont relèvent les raisons à caractère moral, et les raisons non déontiques auxquelles appartiennent les raisons hédoniques. La volonté qu'a Parfit de catégoriser, voire d'opposer, ces deux types de raisons alors que Scanlon se refusait explicitement à distinguer (et par conséquent à hiérarchiser) le bien-être des autres valeurs personnelles<sup>14</sup>, n'est pas anodine. La similarité syntaxique de ce dualisme n'est, en effet, pas sans rappeler celui de Sidgwick bien qu'ils diffèrent sémantiquement (l'opposition point de vue subjectif/point de vue objectif n'est pas réductible à l'opposition raisons non déontiques/raisons déontiques). En réglant la seconde opposition, Parfit peut donc donner l'impression de régler la première.

Avant de revenir sur la comparaison entre raisons déontiques et raisons non déontiques, il est nécessaire de clarifier la procédure de comparaison en jeu au sein de la catégorie des raisons non déontiques selon qu'elles sont relatives à la personne (c'est-à-dire qu'elles sont intéressées) ou non. Bien que les raisons soient objectives, dans le sens où elles sont données par l'objet, elles ne s'appliquent pas nécessairement de façon égale à tous. C'est une implication directe de sa conception de l'hédonisme développée dans la section précédente. Il en généralise la portée dans ce qu'il appelle la thèse du particulier-au-général (*The Any-All Thesis*) selon laquelle « si quelqu'un a une raison personnelle de vouloir que quelque chose arrive, tout le monde a une raison impartiale correspondante, *bien que peut-être de façon beaucoup plus faible*, de vouloir que cette chose arrive également<sup>15</sup> ». Le fait que je reconnaisse que les raisons d'autrui ne soient pas moins fondées que les miennes ne m'engage pas à les traiter comme les miennes. Elles n'ont pas pour moi la force qu'elles ont pour les autres. Si la souffrance<sup>16</sup> constitue toujours une raison de vouloir l'éviter,

13. T. M. SCANLON, *What We Owe to Each Other*, Cambridge, Mass., The Belknap Press, 1998, p. 243.

14. *Ibidem*, p.142.

15. *On What Matters III*, Oxford, Oxford University Press, 2017 (noté désormais *OWM III*), p. 321, je souligne.

16. Notons que la grande majorité des exemples présentés par Parfit évoque la souffrance ou la mort potentielle. C'est le cas notamment (dans *On What Matters*) de *Lift*, de *Tunnel*, de *Sidetrack* et de *Bridge* (qui sont tous des reformulations du célèbre cas du *Trolley*), de celui de la *Nuclear Policy* (inspiré des bombardements d'Hiroshima et Nagasaki), ou encore de ces exemples fictifs aux noms évocateurs : *Child Drowning*, *Shipwreck*, *Lifeboat*, *Fire*, *Earthquakes*, *Desperate Plight*, *Whimsical Despot*, *Murderous Theft*, *Gangster*, *Lesser Evil*, *Depression*, etc. Rares sont les exemples qui justifient la force d'une raison non déontique par l'intensité et/ou la quantité de plaisir

elle n'a pas la même force si elle me concerne, ou concerne mes proches, que si elle concerne des inconnus. Cette considération peut, cependant, être contrebalancée par le nombre des personnes en souffrance, en nous amenant, par exemple, à préférer nous sacrifier pour sauver un nombre plus important de personnes, sans que l'on ne puisse jamais établir un critère de délibération fixe. Les raisons non déontiques, fondées sur la gravité des souffrances ou l'ampleur des plaisirs, sont toujours comparables, mais elles ne le sont qu'*approximativement*.

L'évaluation de la force des raisons déontiques, et par conséquent de leur comparabilité, pose néanmoins plus de problèmes. Les raisons déontiques peuvent être définies comme des devoirs ou des prohibitions de nature non conséquentialiste. Il en existe plusieurs types. Les premières relèvent de principes ayant valeur (universelle ou non) qui se justifient par le fait qu'ils pourraient être rationnellement choisis et/ou acceptés par tous, selon la version kantienne (*the Kantian formula*), ou que personne ne pourrait raisonnablement rejeter, selon la version de Scanlon (*the Scanlon formula*). Mais il peut également s'agir, non de principes, mais de devoirs ou de prohibitions *prima facie*, selon l'expression de David Ross<sup>17</sup>, autrement dit de devoirs auxquels on doit se conformer à moins qu'on n'ait de meilleures raisons de faire autrement. Les raisons déontiques chez Parfit sont, en quelque sorte, un combiné de ces trois approches. Elles sont personnelles mais ayant néanmoins une dimension générale<sup>18</sup>, et peuvent entrer en conflit les unes avec les autres. Parfit croit ainsi que

nous avons des raisons personnelles et impartiales qui ne nous sont pas fournies, ou pas totalement fournies, par les faits relatifs à notre bien-être ou celui des autres. Nous avons des raisons d'essayer de faire différentes choses qui ont de la valeur en elles-mêmes, telles que découvrir la vérité, créer ou préserver la beauté, réaliser plusieurs autres buts bons ou valables. Nous avons aussi des raisons morales à la fois neutres et relatives à la personne, comme les raisons de prendre soin de nos

qu'elle pourrait procurer. Doit-on en conclure que Parfit défend, au moins implicitement, une forme de conséquentialisme négatif? On peut raisonnablement le penser. Dans *Reasons and Persons*, Parfit affirme néanmoins que « le soulagement de la souffrance devrait avoir un poids plus important que celui qui lui est ordinairement donné par les utilitaristes positifs. Mais il ne semble pas plausible qu'il doive avoir priorité absolue, ou même un poids beaucoup plus important. » (*RP*, pp. 344-45).

17. W. D. ROSS, *The Right and the Good*, Oxford, Clarendon Press, 1930.

18. J'utilise ici à dessein le terme de « général » plutôt que celui d'« universel ». Si les raisons ont clairement une dimension universelle chez Kant, elles n'en ont pas chez Scanlon qui accepte un certain pluralisme des raisons lié notamment aux différences de culture. La position intermédiaire de Parfit à cet égard ne me semble pas très claire. Dans l'incertitude, je préfère donc parler de dimension générale des raisons.

propres enfants, et des raisons de faire en sorte que les avantages et les inconvénients soient plus également distribués entre les différentes personnes<sup>19</sup>.

Comment comparer une raison non déontique, telle que la recherche du confort personnel, à une raison déontique, telle que la préservation de la nature ? Selon que l'on raisonne en conséquentialiste ou en déontologue, on aura tendance à donner systématiquement plus de poids aux raisons déontiques ou, au contraire, aux raisons non déontiques, remettant en cause le dépassement de l'indétermination sidgwickienne. Une possibilité pourrait, cependant, consister à évaluer le poids des raisons déontiques en fonction de la valeur de la règle qui les porte, autrement dit, à donner une raison non déontique aux raisons déontiques. C'est ce que fait le conséquentialisme de la règle, auquel la théorie de Parfit semble conduire.

*Pourquoi un conséquentialisme de la règle.*

Parfit est très critique vis-à-vis de l'utilitarisme de l'acte mais il reste à la fois très attaché au conséquentialisme et au principe de l'hédonisme. Cela se traduit par sa défense d'un conséquentialisme de la règle kantien<sup>20</sup> qu'il estime être compatible avec le contractualisme de Scanlon. Ce conséquentialisme kantien de la règle (*Kantian rule consequentialism*) est ainsi défini : « Chacun devrait suivre les principes optimifiques, car ce sont les seuls principes que chacun pourrait rationnellement vouloir devenir des lois universelles<sup>21</sup>. » Combiné au contractualisme de Scanlon, on obtient le principe de la triple théorie, à savoir qu'« une action est mauvaise [*wrong*] lorsque ce type d'action est interdit par un principe qui est optimifique, universellement souhaitable par chacun [*uniquely universally willable*], et qui ne peut être raisonnablement rejeté<sup>22</sup> ». Le principe de la triple théorie par lequel s'achève le premier volume de *On What Matters* est présenté comme l'aboutissement ultime de sa réflexion éthique.

19. D. PARFIT, *OWM III*, p. 321.

20. Il s'agit du produit logique de deux principes universellement acceptés (*UA principles*), l'argument kantien selon lequel « tout le monde devrait suivre les principes dont l'acceptation universelle est rationnellement voulue ou choisie par tous » (principe A) et l'argument conséquentialiste selon lequel « il existe des principes optimifiques universellement acceptés que tout le monde a des raisons impartiales de choisir » (principes C et D, *OWM I*, p. 378). Parfit appelle cela l'Argument kantien pour le conséquentialisme de la règle (*OWM I*, p. 379).

21. *OWM I*, p. 411. Plus exactement, il affirme que nous avons de « fortes raisons » d'accepter ce principe (*OWM I*, 418). Notez l'emploi ici du « ought » qui joue d'une certaine ambiguïté entre l'idée de devoir rationnel (avoir suffisamment de raisons pour) et celle de devoir moral (être dans l'obligation morale de).

22. *OWM I*, p. 413.

L'hypothèse que je voudrais formuler est que la triple théorie est censée conforter la position méta-éthique de Parfit, qui se décrit lui-même comme un adepte du cognitivisme normatif non métaphysique et non naturaliste<sup>23</sup>. À l'instar d'autres réalistes moraux, comme Scanlon, il considère donc qu'il existe des normes morales et que celles-ci ne sont pas réductibles à des faits naturels<sup>24</sup>. Parfit considère que le réalisme moral est fragilisé par l'existence de plusieurs philosophies morales rivales et que la convergence de ces différentes théories est possible sous certaines conditions<sup>25</sup>. La thèse de la convergence des différentes théories éthiques permet de reconnaître la validité de chacune sans remettre en question l'existence objective de normes. Le conséquentialisme de la règle a essentiellement pour objet d'intégrer l'équation kantienne dans un modèle contractualiste, que l'on peut qualifier après Scanlon de contractualisme du bien-être (*welfare contractualism*). Il lui permet également de mettre l'accent sur la dimension hédoniste des raisons que Scanlon tend, selon lui, à minimiser. Reste que le modèle des raisons déontiques et non déontiques développé par Parfit ne peut être présenté comme un conséquentialisme de la règle dans la mesure où l'évaluation de la force des raisons est fonction des personnes auxquelles elles s'imposent. Il ne peut donc pas y avoir de systématisme de la règle dans le modèle de Parfit.

Le conséquentialisme de la règle ne garantit ensuite pas, dans sa version traditionnelle, le respect des raisons déontiques. Le conséquentialisme de la règle juge de la valeur de la règle uniquement en fonction de sa contribution à la maximisation du bien total. Il est bon de dire la vérité car on peut raisonnablement penser que dire la vérité est, en général, une règle *optimifiqu*e. Supposons un instant, cependant, que nous vivions dans une société qui valorise les « *fake news* » et que l'on attache plus d'importance à ce que l'on veut croire qu'à ce qui est vrai. On aurait alors tort, dans ce contexte, de dire la vérité. Cette position n'est, cependant, pas conforme à la pensée de Parfit, pour lequel des raisons déontiques peuvent toujours contre-balancer des raisons non déontiques. Ainsi, explique Parfit, « on peut raisonnablement penser que, dans de nombreux cas, il serait mal en soi de traiter les gens d'une certaine façon, telle que les tromper ou les contraindre, violer les promesses qu'on leur a faites. Le mal intrinsèque de telles actions surpasserait souvent le bien de leurs conséquences<sup>26</sup> ». Il ne peut donc exister pour Parfit de règles optimifiques fondamentalement contraires aux raisons déontiques (même si des conflits

23. D. PARFIT, *On What Matters II*, Oxford, Oxford University Press, 2011 (noté désormais *OWM II*), p. 486.

24. Voir l'article de Laurent Jaffro dans la présente livraison.

25. *OWM II*, pp. 549-565.

26. *OWM III*, p. 421.



occasionnels sont possibles). Or, rien, en théorie, n'interdit au conséquentialisme de la règle de promouvoir une règle « immorale » si celle-ci est optimisatrice.

Pour répondre à cette objection, Parfit propose donc de ne pas circonscrire le conséquentialisme à l'évaluation des effets de la règle mais de l'appliquer également à ceux des motifs qui lui sont sous-jacents. C'est ce qu'il appelle le « conséquentialisme de la règle et de la motivation » (*rule and motive consequentialism*)<sup>27</sup>. Si l'impartialité peut se justifier par les effets de la règle (dans la mesure où une égalité de traitement maximise l'utilité totale), la règle selon laquelle il est bon de préférer ses proches aux étrangers se justifie, elle, par l'utilité sociale de la motivation qui l'anime. Un monde dans lequel on n'attacherait pas plus d'importance à ses enfants ou ses parents qu'à de parfaits inconnus ne serait pas un monde heureux. L'optimificité de la règle repose donc conjointement sur les effets de la généralisation de l'action et sur celle de sa motivation. Les uns tempèrent (ou confortent) les autres. À utilité sociale égale, la bonté des motivations d'une règle renforce la raison de s'y soumettre alors que la malveillance inhérente (*intrinsic badness*) l'affaiblit.

Le conséquentialisme de la règle et de la motivation présente un autre avantage par rapport au conséquentialisme de l'acte : il permet de mieux appréhender l'imperceptibilité des effets de certaines actions dont l'action répétée est responsable d'immenses dommages. C'est le cas notamment des actions visant à soulager la pauvreté dans le monde ou des actions visant à limiter la pollution et préserver la biodiversité. Les effets de la règle, ajoutés à ceux de la motivation, donnent aux individus de fortes raisons d'agir en ce sens. Mais ces actions, dont la raison est, je le rappelle, donnée par l'objet (ici le soulagement de la pauvreté et la sauvegarde de la planète), ont-elles véritablement besoin d'être justifiées par les conséquences de leur généralisation ? Cela ne me semble pas évident. L'action qui consiste à aider son prochain est bonne, qu'elle soit suivie par tous ou non. En s'écartant de son cadre théorique de départ, Parfit remet en cause l'objectivité même de sa théorie. Le fondamentalisme des raisons répond déjà bien au problème de l'imperceptibilité des effets des actions individuelles. La question des bonnes actions motivées par un mauvais sentiment (ou inversement) ne se pose pas puisque la raison de l'action est donnée par son objet. La motivation personnelle de l'agent ne rentre pas plus en ligne de compte que son caractère moral.

Pourquoi, alors, conclure le troisième et dernier volume de *On What Matters* par une défense du conséquentialisme de la règle ? La principale raison, on l'a

27. *OWM III*, p. 418.

souligné, est d'ordre méta-éthique. La thèse de la convergence des philosophies morales peut apparaître comme un argument en faveur du réalisme moral. La triple théorie, union sacrée entre l'utilitarisme, la déontologie et le contractualisme, repose sur un syllogisme, selon lequel chacune de ces théories éthiques étant réductible à une forme de conséquentialisme de la règle, ces trois théories seraient donc une et seule même théorie. Ce raisonnement me paraît fragile et ses présupposés contestables. Dans l'espace du présent article, je peux seulement souligner qu'il dépend essentiellement de l'adoption du conséquentialisme de la règle. Pourtant il y a une autre inspiration dans Parfit. Il a parfois été reproché à Parfit de ne pas clairement s'exprimer sur ce qu'il considérait être « important » et donc de ne traiter que très indirectement du sujet de son ouvrage (*what matters*). La fin du troisième volume se distingue radicalement du reste de l'ouvrage puisque Parfit prend la liberté de souligner l'importance du don altruiste qu'il estime devoir être une règle morale personnelle, rejoignant ainsi les partisans de l'altruisme effectif<sup>28</sup>, révélant ainsi peut-être la véritable dimension utilitariste (et non plus simplement conséquentialiste) de sa pensée.

#### *Conclusion.*

L'ambition de cet article était d'éclaircir les relations entre la théorie des raisons que Parfit développe en s'inspirant de Scanlon et le conséquentialisme de la règle par lequel se conclut le troisième volume de *On What Matters*. J'ai tenté de montrer que l'adoption d'un conséquentialisme de la règle et de la motivation ne présentait aucun avantage par rapport au modèle des raisons déontiques et non déontiques développé par Parfit et que celle-ci devait avant tout servir la thèse de la convergence des théories éthiques. Un point mérite à ce stade d'être prolongé pour compléter cette réflexion. Le premier a trait aux raisons éthiques (plutôt que méta-éthiques) qui ont conduit Parfit à défendre l'idée d'une triple théorie. Une hypothèse possible serait, en effet, que cela lui permet de défendre indirectement une position utilitariste qu'il refuse d'adopter ouvertement. Mais cela exige d'analyser beaucoup plus en détail la relation existant entre conséquentialisme et utilitarisme (de la règle), ainsi que la différence théorique entre la thèse de la convergence des théories éthiques et celle de la triple théorie éthique, ce qu'il ne m'était pas possible de faire dans cet

28. Parfit conseille de donner dix pour cent de ses revenus à des œuvres de charité (*OWM III*, p. 436).

article. La position de Parfit à l'égard de l'utilitarisme, que l'on croyait clarifiée dans *On What Matters*, n'a donc pas fini de susciter des débats.

Christophe SALVAT  
*CNRS (Centre Gilles-Gaston-Granger)  
& Aix-Marseille Université*

## L'ÉTHIQUE À L'ÉPREUVE DE LA GESTATION POUR AUTRUI (GPA)

CHRISTOPHE SALVAT

161

Le 17 janvier 2018 se sont ouverts les états-généraux sur la bioéthique. Parmi les questions abordées, la Procréation Médicale Assistée (PMA) et la Gestation pour Autrui (GPA). Cette dernière, en particulier, fait l'objet de très vifs débats dans la presse. On lui reproche notamment de réifier ou de marchandiser le corps humain<sup>1</sup>. D'autres avancent, au contraire, que la GPA – correctement encadrée – constituerait une avancée sociale importante, reconnaissant ainsi la diversité des modèles familiaux<sup>2</sup>. Certains y voient l'apanage du néo-libéralisme, d'autres – au contraire – une forme de collaboration sociale<sup>3</sup>.

L'objet de cet article est de s'interroger sur la légitimité morale d'une loi autorisant la GPA en France aujourd'hui. Notre questionnement sera strictement éthique. Au-delà de toute conclusion, si conclusion il y a, l'enjeu est également d'évaluer la capacité de nos théories éthiques à répondre à des enjeux inédits tels que celui-ci.

On reconnaît généralement trois grandes approches en philosophie morale, la déontologie, le constructivisme et le conséquentialisme. L'argument kantien, strictement parlant, est celui qui est le plus utilisé par les opposants à la GPA, mais – ainsi que nous le montrons – le cadre théorique kantien n'est pas vraiment

- 
1. Sylviane Agacinsky, Ana-Luana Stoicea-Deram, Martine Segalen, « Lettre ouverte à Emmanuel Macron sur la GPA », *Le Figaro*, 18 avril, 2017. Voir aussi, Sylviane Agacinski, *Corps en Miettes*, Paris, Flammarion, 2009.
  2. Collectif, « On ne peut plus ignorer les enfants nés par GPA », *Le Monde*, 16 janvier 2018 ; Christophe Salvat, « Pour une réglementation éthique de la GPA en France », *Libération*, 24 mai 2017.
  3. Marianne Durano, « PMA-GPA : le business plan est prêt », *Le Figaro*, 23 janvier 2018 ; Irène Théry, « Le portrait-robot du français individualiste et sans valeurs morales ne correspond pas à la réalité », *L'Opinion*, 12 janvier 2018.

adapté au cas de la procréation assistée. L'approche constructiviste, d'inspiration rawlsienne ou scanlonienne, a de plus fortes possibilités théoriques mais ne permet pas, croyons-nous, de prendre véritablement position. Elle bute notamment sur un obstacle majeur, soulevé par Michael Sandel<sup>1</sup>, à savoir qu'elle implique nécessairement de faire des hypothèses quant aux valeurs et préférences des individus derrière le voile d'ignorance. Quant à la théorie conséquentialiste, et à sa forme traditionnelle, l'utilitarisme, elle semble *a priori* donner raison aux partisans de la GPA mais elle soulève un certain nombre de problèmes en fonction de la théorie de la valeur qu'elle adopte. C'est la raison pour laquelle je termine en discutant une quatrième et dernière approche, à mi-chemin entre déontologie kantienne, constructivisme et conséquentialisme, à savoir l'utilitarisme des raisons de Derek Parfit.

162

#### L'ARGUMENT KANTIEN DE LA DIGNITÉ HUMAINE

L'impératif pratique « Agis de telle façon que tu traites l'humanité, aussi bien dans ta personne que dans celle d'autrui, toujours en même temps comme une fin, et jamais simplement comme un moyen » est souvent avancé pour justifier l'immoralité de la GPA. Il fait du devoir envers soi-même l'un des principes fondateurs de la philosophie morale kantienne. Chacun doit, selon Kant, traiter sa propre personne avec le même égard dû à toute autre personne. Cela signifie, entre autres choses, que l'on a pas plus le droit de se suicider que de commettre un meurtre, de se mutiler que de torturer autrui, ou dans le cas qui nous intéresse ici, de se dégrader moralement que de dégrader les autres. Cela serait contraire au principe de dignité humaine. Or, affirment les opposants de la GPA, enfanter pour autrui est un acte dégradant.

La GPA, pour Sylviane Agacinski, « porte atteinte à l'intégrité physique de la mère, en engageant profondément sa vie organique et psychique sans nécessité thérapeutique. Elle touche aussi à la dignité de l'enfant, sur la tête de qui un contrat a été passé. Car c'est bien l'enfant qui, in fine, fait l'objet de la transaction »<sup>2</sup>. Le même argument est cité par le Comité Consultatif National d'Éthique pour les Sciences de la Vie et de la Santé dans son avis 110 relatif aux problèmes soulevés par la Gestation pour Autrui en 2010 :

1. Michael Sandel, *Liberalism and the Limits of Justice*, Cambridge, Cambridge University Press, 1982.

2. Sylviane Agacinski, « La gestation pour autrui est un commerce dégradant », *La Croix*, 16 avril 2009.

La dignité, qualité intrinsèque de l'humanité, interdit, dans une conception kantienne, de considérer l'homme seulement comme un moyen et de lui conférer un prix. (...) la GPA représente une instrumentalisation des corps des femmes et aboutit à considérer l'enfant comme une marchandise, de sorte que cette pratique est d'une manière irréductible contraire au respect de la dignité de la personne humaine. (...) Même si l'on peut admettre que certaines gestatrices agissent dans le cadre d'une liberté effective, il reste qu'on peut ne pas souhaiter que la société accepte ce qui apparaît comme une aliénation, fût-elle volontaire.<sup>1</sup>

Le contrat (moral et/ou juridique selon les cas) relatif à la GPA constitue-t-il une forme de dégradation volontaire ? La prostitution l'est certainement, du moins pour Kant. Mais, rappelons néanmoins, que ce n'est pas sa nature contractuelle et monétaire qui rend la prostitution indécente, mais le fait même qu'il s'agisse d'une activité sexuelle hors mariage et déconnectée de toute fin reproductive. Or, toutes les activités sexuelles qui n'ont pas pour objet la reproduction (dans le cadre légal du mariage), c'est-à-dire la copulation 'de loisir', la masturbation et les pratiques homosexuelles, sont considérées par Kant comme dégradantes. L'activité sexuelle consiste à user, dans le but de se donner du plaisir, du corps d'autrui ou de son propre corps (dans le cas de la masturbation). Elle fait donc de la personne humaine un simple moyen, et est, à ce titre, contraire à la dignité humaine. Elle est donc condamnable, sauf lorsque celle-ci prend place dans le cadre du mariage entre un homme et une femme, mariage que Kant définit comme un contrat ayant pour objet « la liaison de deux personnes de sexe différent en vue de la possession réciproque, pour toute la durée de la vie, de leurs qualités sexuelles propres »<sup>2</sup>. Ce contrat, fondé sur le consentement des deux parties, légitime le droit d'user de l'autre, et de son corps, comme l'on userait d'un objet. Comme une personne n'est cependant pas assimilable à un objet, Kant parle de droit 'personnel d'espèce réelle' plutôt que de droit réel.

163

La Gestation pour Autrui permet d'enfanter, dans un cadre légal, en se passant de toute relation sexuelle. Cela aurait-il déplu à Kant ? Je n'en suis pas sûr. Peut-on raisonnablement recourir à ses arguments pour la condamner ? C'est discutable. Si tel est le cas, sa nature contractuelle ne peut être seule en cause. Mais, à défaut d'arguments plus spécifiques de la part de Kant, qui n'avait évidemment pas les moyens d'anticiper la possibilité d'une reproduction asexuée, on est forcé de conclure que la théorie kantienne ne permet ni de justifier ni de condamner la GPA.

1. Comité Consultatif National d'Éthique pour les Sciences de la Vie et de la Santé, « Problèmes éthiques soulevés par la Gestation pour Autrui (GPA) », *AVIS n°110*, 2010, p.14.

2. Kant, *Doctrine du droit*, trad. A. Renaud, Paris, GF-Flammarion, 1994, p. 77-78.

## GPA ET VOILE D'IGNORANCE

164

Les théories constructivistes d'inspiration kantienne pourraient, elles, nous fournir plus d'arguments. Celles-ci, à l'instar de la théorie de la justice de John Rawls<sup>1</sup> ou de la philosophie morale de Thomas Scanlon<sup>2</sup>, font du consentement rationnel et impartial la source de légitimité toute norme de justice (ou de morale). Le critère d'impartialité est assuré, chez Rawls, par l'expérience de pensée dite du voile d'ignorance, derrière lequel les individus dépourvus de leur identité, de leurs préférences et de leur valeur, sont amenés à se prononcer sur ce qu'ils considèrent être acceptable (ou, dans le cas de Scanlon, sur ce qu'ils ne peuvent raisonnablement considérer comme inacceptable). Le critère de rationalité est assuré par une connaissance supposée parfaite des individus et une conception objective de leur bien. L'intérêt (certains diront la limite) de cette approche est d'être totalement décontextualisée, y compris d'un point de vue temporel. Derrière le voile de l'ignorance, les individus ne savent pas à quelle génération ils appartiennent ou appartiendraient s'ils venaient à exister. Dans le cas qui nous occupe, cela revient à demander à des personnes, qui n'existent pas encore mais dont la probabilité d'exister est *de facto* augmentée par la GPA, si elles considèrent cette pratique comme acceptable. Il y a de fortes raisons de penser qu'elles ne s'y opposeront pas. S'agit-il pour autant d'un bon argument ? Je ne le crois pas. Tout d'abord, ainsi que Michael Sandel l'a fait remarquer, il est inconcevable de demander à des individus privés de leur identité, de leurs préférences et de leurs valeurs de se prononcer sur ce qu'il leur semble être bon pour eux<sup>3</sup>. Ensuite, le fait qu'ils aient tous individuellement intérêt à quelque chose ne signifie pas nécessairement que cela soit dans leur intérêt. Prenez le cas de la population future. N'étant pas nés, nous avons tous individuellement intérêt à naître. Nous avons donc tous intérêt à approuver toutes les mesures visant à augmenter le nombre de naissances (puisque celle-ci augmente *notre* probabilité de naître). Mais, nous n'avons pas pour autant tous intérêt à vivre dans un monde surpeuplé. Ce qui est serait rationnel pour moi, c'est de faire en sorte que *ma* probabilité de naître augmente sans que cela n'affecte celle des autres. Mais cela, la théorie contractualiste est incapable de l'appréhender. J'en conclus, donc, que bien que le contractualisme puisse justifier la GPA, cette justification est problématique<sup>4</sup>.

1. John Rawls, *Théorie de la justice*, Paris, Editions du Seuil, 1997.

2. Thomas M. Scanlon, *What Do We Owe to Each Other ?*, Cambridge (MA), Harvard University Press, 2000.

3. Michael Sandel, *Liberalism and the Limits of Justice*, Cambridge, University Press, Cambridge, 1982.

4. Ces arguments ne sont pas exhaustifs. On pourrait également citer celui de Carol Pateman, selon laquelle le consentement n'est pas un critère de légitimité lorsque le contrat lui-même est l'expression d'un système

## LES APPROCHES CONSÉQUENTIALISTES

J'en viens donc à la troisième et dernière approche éthique, à savoir le conséquentialisme. Il en existe trois grands types selon la théorie de la valeur que l'on adopte, à savoir l'hédonisme ou théorie des états mentaux, la théorie du succès ou des états du monde, et la théorie des listes objectives<sup>1</sup>. Les théories hédonistes ont *a priori* des raisons de soutenir toutes les lois visant à réduire l'infertilité et augmenter la natalité dans la mesure où chaque nouvelle vie contribue à maximiser l'utilité totale. C'est le cas lorsque la naissance est désirée et qu'elle procure une somme nette de plaisirs aux parents et à ceux qui naissent, mais également à toutes les générations issues de ce premier enfantement. Compte-tenu de la souffrance éprouvée par les nombreux couples hétérosexuels ou homosexuels incapables de procréer dans une société qui valorise fortement la cellule familiale (bien que sur un autre modèle que les générations passées), il n'est pas déraisonnable de penser que la GPA augmentera significativement le bien être des couples infertiles<sup>2</sup>. Il n'y pas de raison de penser, en revanche, que le recours à la GPA réduira significativement le bien être de ceux qui n'ont pas besoin de recourir à la GPA pour enfanter (bien que cette hypothèse ne puisse pas être totalement exclue). Quant à l'utilité nette des enfants à venir, elle est naturellement toujours délicate à évaluer, et dépend, notamment de leur insertion sociale. Les études menées, notamment en Angleterre où l'adoption par les couples homosexuels et la Gestion pour Autrui est autorisée, concluent que le bien-être des enfants élevés dans des familles homoparentales n'est pas très différent de celui des enfants élevés par des familles hétérosexuelles, voire qu'il est parfois plus élevé<sup>3</sup>. On peut donc raisonnablement penser que la GPA contribuera à augmenter l'utilité totale des individus concernés. D'un strict point de vue utilitariste, le bonheur des parents et des enfants ne compte cependant pratiquement pour rien au regard de celui

d'exploitation sexuelle, comme c'est le cas, selon elle, des contrats de mariage, des contrats de prostitution ... et des contrats de mères porteuses. Carol Pateman, *The Sexual Contract*, Stanford, Stanford University Press.

1. Derek Parfit, *Reasons and Persons*, Clarendon Press, Oxford, p.4.
2. Dans quelle mesure peut-on affirmer que l'utilité nette gagnée sera réellement significative ? Au regard du faible nombre de couples concernés, on peut légitimement se poser la question. Néanmoins la prise en compte du degré de souffrance, associé à l'impossibilité de fonder une famille, les attentes souvent très longues et les épreuves passées, il est raisonnable de penser que malgré le faible nombre de personnes concernées le gain d'utilité sera suffisant pour être notable. Ce point mérite attention car bien que les approches utilitaristes traditionnelles considèrent que toute augmentation d'utilité, aussi minime soit-elle, justifie l'action, certains développements récents, notamment ceux de Parfit, remettent en question ce principe, considérant que seuls des gains d'utilité substantiels peuvent constituer une raison suffisante.
3. Laura Mellish, Sarah Jennings, Fiona Tasker, Michael Lamb and Susan Golombok, « Gay, Lesbian and Heterosexual Adoptive Families Family relationships, child adjustment and adopters' experiences », *Report of the British Association of Adoption and Fostering (BAAF)*, 2013.



de toutes les générations futures affectées par ces naissances. Or, l'évaluation de ces utilités (ou désutilités) futures est très incertaine et repose sur des hypothèses extrêmement hasardeuses. Une hypothèse pessimiste sur le sort de l'humanité suffirait à remettre en cause la légitimité de la GPA (ou de toute autre mesure populationniste). L'utilitarisme classique ne permet donc pas de se prononcer avec certitude sur la légitimité de cette pratique. Dans la pratique, celui-ci a, cependant, largement laissé la place à l'utilitarisme dit des préférences, qui repose sur l'évaluation des états du monde plutôt que sur celles des états mentaux des individus.

166

La théorie du succès peut-elle réussir là où l'hédonisme a échoué ? Rien n'est moins sûr. A première vue, la théorie des états du monde a l'avantage de ne pas être dépendante de spéculations sur le ressenti des autres et, *a fortiori*, de ceux qui n'existent pas encore. Bien que les préférences individuelles soient, comme leur nom l'indique, propres à chacun, on peut néanmoins supposer que les individus ont, dans leur très grande majorité, une préférence à vivre, y compris dans des circonstances qui leur sont défavorables. Cela évacue donc l'un des principaux problèmes soulevés par l'hédonisme. Néanmoins, il est toujours possible de supposer que certaines personnes aient une préférence à être né 'naturellement' ou dans un couple hétérosexuel. Comment alors prendre en compte cette frustration ? Est-il préférable pour une personne ne souhaitant pas être née par GPA de ne pas exister ? Comment, ensuite, arbitrer entre la préférence des parents d'intention d'avoir un enfant par GPA et celle de l'enfant de ne pas naître par GPA ? Plus fondamentalement, et contrairement à l'hédonisme, la théorie des préférences donne une place exagérément importante à tous ceux (et ils sont nombreux), qui ne sont pas concernés par cette pratique, qui n'ont pas de difficulté à enfanter et dont le bien-être ne dépend pas de l'existence (ou non) de cette mesure, mais qui néanmoins préféreraient que les autres ne se reproduisent pas par GPA. Cela me paraît moralement inacceptable.

Reste les théories des listes objectives, qui sont à l'heure actuelle très peu utilisées. Le premier à avoir développé une théorie objective du bien dans une perspective conséquentialiste est G.E. Moore, premier théoricien de ce que l'on a appelé l'utilitarisme idéal<sup>1</sup>. Elles souffrent de deux gros handicaps. Le premier est d'ordre pratique. Elles ne sont pas quantifiables ou ne le sont que très diffi-

1. Sur ce mouvement peu documenté voir Anthony Skelton, « Ideal Utilitarianism: Rashdall and Moore », In T. Hurka, editor, *Underivative Duty: British Moral Philosophers from Sidgwick to Ewing*, Oxford University Press, Oxford, 2011.

cilement, ce qui les rend *de facto* inopérantes. Le second est d'ordre théorique. Cette conception de la valeur, qui est aujourd'hui le plus souvent utilisée par des théories critiques de l'utilitarisme (notamment par Rawls et par Sen), a cependant souvent du mal à justifier le choix des éléments qu'elle considère 'objectivement' bons. Elle a néanmoins l'avantage de ne pas conclure qu'une vie est bonne ou qu'elle vaut la peine d'être vécue simplement parce que la personne ne souffre pas d'être en vie ou qu'elle a une préférence à rester en vie. D'autres éléments, tels que les libertés fondamentales, un revenu suffisant, un accès à l'éducation etc., rentrent en ligne de compte. Contrairement aux théories précédentes, la théorie des listes objectives n'a pas vocation à répondre à la question de la légitimité de la GPA, considérée en général, mais elle peut apporter un certain nombre d'arguments en sa faveur ou en sa défaveur dans un contexte particulier, voire même établir des conditions minimales d'application pour que celle-ci soit acceptable. On peut, par exemple, raisonnablement penser que la GPA, telle qu'elle se pratique en Inde ou dans un certain nombre de pays du Sud, n'est pas conforme aux principes de libertés et de droits sociaux fondamentaux mis en avant par John Rawls ou Amartya Sen<sup>1</sup>. En revanche, la théorie des listes objectives ne semble pas disqualifier une pratique éthique de la GPA, telle qu'elle est appliquée au Royaume-Uni.

167

La capacité du conséquentialisme à répondre à la question de la légitimité morale de la GPA dépend, en grande partie de la théorie de la valeur sur laquelle cette théorie repose. La théorie de l'hédonisme est la plus exclusive de toutes puisqu'elle considère le plaisir comme seule motivation de l'action. La théorie des préférences est, au contraire, trop inclusive puisqu'elle inclut toutes les inclinations humaines, y compris les plus asociales et les immorales. La théorie des listes objectives, enfin, permet de mieux discriminer nos fins, bien qu'elle ait du mal à justifier ses choix. On pourrait, cependant, lui reprocher de ne pas accorder suffisamment d'importance au plaisir des individus. On peut critiquer l'exclusivité des raisons que s'arrogent les théoriciens de l'hédonisme psychologique, on peut également regretter son fondement subjectif (qui défie souvent les règles du bon sens), mais on peut difficilement exclure toute motivation hédoniste des agents. C'est la raison pour laquelle je considère que la tentative de Parfit de concilier hédonisme et objectivité des valeurs est la bienvenue.

1. Amartya Sen, *L'idée de justice*, Paris, Flammarion, 2010. Sur l'application de la théorie des capacités à la GPA en Inde, voir Sheela Saravanan, « Global Justice, capabilities approach and commercial surrogacy in India », *Medicine, Health Care and Philosophy*, August 2015, Volume 18, Issue 3, pp 295–307.

## UNE INTERPRÉTATION PARFITIENNE DE LA GPA

Dans *On What Matters*, Parfit distingue deux types de raisons d'agir : les raisons déontiques, c'est-à-dire les obligations morales du type des *prima facie duties* de Ross<sup>1</sup>, et les raisons non-déontiques, ou conséquentialistes, au sein desquelles il range les raisons hédonistes<sup>2</sup>. Contrairement à l'approche utilitariste traditionnelle, l'hédonisme constitue, pour Parfit, une raison objective, c'est-à-dire que la souffrance constitue toujours pour l'agent une raison d'agir (en vue de l'éviter), même si celui-ci n'est pas concerné par celle-ci. Parfit admet simplement que le sujet n'étant pas le bénéficiaire de l'action, cette raison a moins de force pour lui qu'elle en aurait eu s'il l'avait été. Il est donc possible de comparer, au moins approximativement, les raisons qui sont personnelles de celles qui ne le sont pas.

**168** Dans le cas de la GPA, qui nous occupe ici, cela signifie qu'il est théoriquement possible, non seulement, de reconnaître la coexistence de valeurs morales de type déontologique et de raisons d'ordre conséquentialistes (qu'elles soient personnelles ou non), mais également aussi de comparer leur force relative. En d'autres mots, cela signifie qu'on peut à la fois considérer que la GPA transgresse l'interdit moral de la marchandisation des corps et qu'elle contribue à l'amélioration du bien-être humain. Dans cette configuration nouvelle, aucune de ces deux raisons ne peut être, à elle seule, décisive. La légitimité finale de l'action dépend de la force relative de ces raisons. Le débat n'est certes pas tranché, car l'évaluation des forces relatives des raisons déontiques et non déontiques pose de vraies difficultés, mais il a néanmoins significativement avancé. Parfit n'aborde malheureusement pas les questions d'éthique appliquée. Son discours reste donc entièrement théorique. Le quatrième et dernier volume de *On What Matters*, qui ne verra jamais le jour, devait être plus explicite sur ce que Parfit considérait comme objectivement 'bon'. A défaut d'un positionnement explicite de sa part, on peut néanmoins suggérer plusieurs pistes de réflexions.

La première question à laquelle nous devons répondre est la suivante : quelle force donner aux raisons déontiques souvent avancées dans le débat ? Celles-ci ont quatre dimensions : une dimension religieuse, une dimension idéologique, une dimension sociale et une dimension morale. Les principales religions pratiquées en France sont opposées à la GPA, qui la considère contraire à la parole de Dieu. Cette interdiction est, pour eux, absolue. Mais la France étant un pays

1. William D. Ross, *The Right and the Good*, Oxford, Clarendon Press, 1930.

2. Derek Parfit, *On What Matters*, Oxford, Oxford University Press, 2011 (vols. 1&2) et 2017 (vol.3).

laïc, la croyance religieuse ne constitue pas une raison suffisante pour autoriser ou interdire une pratique. C'est ainsi que, par exemple, le mariage homosexuel a été autorisé en France en 2013 alors que celui-ci n'est pas reconnu par l'Église. La dimension idéologique du débat sur la GPA est très importante. Je voudrais suggérer, cependant, qu'elle ne peut pas être décisive. Par raison idéologique, on entend les raisons ayant trait aux valeurs que l'on souhaite voir reconnaître dans une société. Il existe deux principaux types de modèles de société : celles qui sont construites sur les notions de responsabilité individuelle, de mérite et de marché – les sociétés dites libérales – et celles qui font de la solidarité et de l'égalité leurs valeurs cardinales, et que l'on associe en général au modèle socialiste. Les opposants à la GPA condamnent ce qu'ils considèrent être le développement ultime du néo-libéralisme, à savoir le droit de vendre ou de louer son corps. Ses partisans mettent, au contraire, l'accent sur la motivation altruiste des mères porteuses et l'égalité qu'elle permet d'instaurer entre couples fertiles et infertiles, couples homosexuels et couples hétérosexuels. Les deux arguments sont parfaitement recevables. Le fait, cependant, qu'aucun de ces deux systèmes ne soit objectivement meilleur que l'autre, et que ni l'un ni l'autre ne fasse l'objet d'un consensus social, ne permet pas de faire de cet argument une raison décisive de soutenir ou de s'opposer à la GPA en France.

La dimension sociale des raisons déontiques a essentiellement trait au statut de la femme dans la société. La GPA constituerait une forme d'aliénation et d'abaissement des femmes dont le statut social (et économique) serait réduit à leur fonction reproductive. A l'instar des prostituées qui porteraient atteinte à toute la condition féminine (voire humaine) en faisant commerce de leurs attributs sexuels (et qui, pour cette raison, sont ostracisées par cette même société qui les exploite), les mères porteuses porteraient atteinte à l'image de la femme, et s'exposeraient à la même sanction sociale. L'argument est valable dans le cas où certaines femmes feraient profession de la gestation. Ce n'est pas ainsi que la GPA est envisagée en France, fort heureusement. Lorsque celle-ci est libre, occasionnelle et indemnisée plutôt que rémunérée, il n'y a pas de raison de penser que le statut social de la femme soit altéré. J'ai déjà évoqué plus haut l'impératif moral, généralement réduit à une lecture très contestable de Kant. Il n'existe pas à ce jour de philosophie morale de type déontologique qui puisse valablement justifier une interdiction *inconditionnelle* de la GPA. Les arguments dits kantien qui nous sont régulièrement présentés dans les médias masquent en fait des raisons

de nature idéologique et/ou religieuse, qui – ainsi qu'on l'a vu – ne peuvent à eux seuls être décisifs.

170 Les raisons non-déontiques ont essentiellement une dimension hédoniste, mais il est également possible de prendre en compte des conséquences non strictement hédonistes de la GPA, telles que les mutations du système économique ou social (les mères porteuses délaissant leurs emplois, la baisse des demandes d'adoption, etc.). Dans la mesure où nous ne discutons ici que d'une proposition de loi visant à instituer une pratique éthique et encadrée de la GPA, les effets économiques de cette pratique seront minimes. Au vu du fort déséquilibre entre le nombre d'enfants proposés à l'adoption et le nombre de couples en attente d'un enfant, on peut penser que l'autorisation de la GPA ne se fera pas au détriment des adoptions. Mettons donc les raisons non-déontiques et non hédonistes, au moins provisoirement, de côté. Reste donc ce que l'on pourrait appeler *l'empreinte hédoniste* de la GPA. Elle est globalement positive, très positive même. D'un côté, un grand nombre de couples privés de la joie de fonder une famille pourront désormais y accéder, de l'autre l'expérience altruiste des mères porteuses qui, lorsque la pratique est bien encadrée, continuent d'être en relation avec les enfants à qui elles ont donné naissance. Mais l'impact le plus important est naturellement celui constitué par l'existence rendue possible de ces enfants et de tous les descendants que ceux-ci pourront avoir. Et puisque, pour Parfit, l'hédonisme est une raison objective, aussi bien personnelle qu'impersonnelle, le bonheur de toutes ces personnes constitue pour tous les autres une raison (bien que d'intensité plus faible) de l'approuver.

## CONCLUSIONS

En fin de compte, on peut conclure, dans une perspective parfitienne, que la force des raisons non-déontiques l'emporte très largement sur les raisons déontiques. Cela suffit-il pour clore le débat et conclure à la légitimité de la GPA éthique en France ? Non, pas tout à fait, car un certain nombre de questions restent, je crois, en suspens. J'ai montré qu'aucune raison déontique n'avait assez de force pour emporter la décision dans un sens ou dans l'autre. Cela me semble assez juste dans l'état actuel du débat philosophique. Mais on pourrait m'opposer que la faiblesse des raisons déontiques morales tient davantage à la faiblesse du débat philosophique qu'à la nature des arguments en question. Cela est vrai. Il est fort possible qu'une véritable théorie déontologique, d'inspiration kantienne ou

non, se développe dans un futur proche et emporte le débat dans un sens ou dans l'autre. Toutes les raisons déontiques morales ne sont, en effet, pas par nature hostiles à la GPA. Il peut ainsi tout à fait exister des raisons déontiques morales non kantienne qui plaideraient en faveur de la GPA. Symétriquement, il existe des raisons non déontiques qui pourraient tout à fait plaider en sa défaveur. J'en ai cité deux dans mon argumentaire, mais il y a en sans doute d'autres. Enfin le fondement hédoniste des raisons déontiques, qui lui donnent toute force, n'est pas sans poser problèmes. Inclure dans le calcul d'utilité le plaisir net des enfants à naître, ainsi que celle de leur descendance, conduit invariablement à soutenir toute mesure populationniste quelle qu'elle soit. Il y a peu de raisons de penser que l'autorisation de la GPA en France nous conduise à une situation du type de la Conclusion Répugnante de Parfit<sup>1</sup>. Son effet sur la population sera marginal et il n'y aucune raison de penser qu'il s'accompagnera d'une réduction de l'utilité moyenne des habitants. Néanmoins on est en droit de s'inquiéter de la force que donne toute mesure populationniste aux raisons non-déontiques, force qui pourrait théoriquement nous amener un jour à accepter des pratiques moins librement consenties que ne le serait la GPA éthique. Pour terminer, et dans la continuité du point précédent, je voudrais signaler ce qui paraît être l'une des plus grosses difficultés à laquelle doit faire face toute philosophie morale à l'égard de la GPA, à savoir celle relative à sa normativité. A supposer qu'une théorie éthique, telle que celle que je viens de présenter, parvienne à démontrer la légitimité morale de la GPA, dans quelle mesure celle-ci deviendrait-elle non seulement moralement acceptable mais, également, moralement contraignante ? Ne sommes-nous pas tous moralement tenus de « bien » agir ? Comment justifier la nature surrogatoire de la GPA ? Toutes ces questions méritent, il me semble, d'être réellement débattues par les philosophes avant que le débat ne soit confisqué, parfois au nom même de la philosophie, par ceux qui ne défendent, en réalité, qu'une vision idéologique.

---

1. « The Repugnant Conclusion » est formulée ainsi par Parfit « For any possible of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living » Derek Parfit, *Reasons and Persons*, *op.cit.*, p. 388. Je reprends ici la traduction de Pascal Engel. <http://lafrance-byzantine.blogspot.fr/search?q=parfit>





**Æconomia**

History, Methodology, Philosophy

5-4 | 2015

Varia / Économie et littérature

---

## Parfit, l'égoïsme rationnel et la question de l'identité personnelle

*Parfit, Rational Egoism and the Issue of Personal Identity*

Christophe Salvat

---



### Édition électronique

URL : <http://journals.openedition.org/oeconomia/2175>

DOI : 10.4000/oeconomia.2175

ISSN : 2269-8450

### Éditeur

Association Æconomia

### Édition imprimée

Date de publication : 1 décembre 2015

Pagination : 437-460

ISSN : 2113-5207

Ce document vous est offert par Aix-Marseille Université (AMU)



### Référence électronique

Christophe Salvat, « Parfit, l'égoïsme rationnel et la question de l'identité personnelle », *Æconomia* [En ligne], 5-4 | 2015, mis en ligne le 01 décembre 2015, consulté le 13 juin 2020. URL : <http://journals.openedition.org/oeconomia/2175> ; DOI : <https://doi.org/10.4000/oeconomia.2175>

---



Les contenus d'*Æconomia* sont mis à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.



# Parfit, l'égoïsme rationnel et la question de l'identité personnelle

Christophe Salvat\*

L'objet du présent article est d'illustrer l'importance de la philosophie en économie en montrant que notre conception de la rationalité individuelle est étroitement liée à notre conception de l'identité personnelle. Au vu de l'importante littérature sur le sujet, j'ai choisi de me limiter à la théorie du philosophe Derek Parfit, telle qu'elle est présentée dans son ouvrage de référence *Reasons and Persons* (1984). Son approche de l'identité, bien que d'inspiration lockéenne, est très originale et a suscité de nombreux débats en philosophie. Elle reste, cependant, encore assez méconnue en économie, en France notamment, malgré l'intérêt croissant que suscite la philosophie économique. Une première partie de cet article s'attache à présenter l'approche réductionniste de Parfit. La deuxième partie et la troisième partie étudie l'impact de cette approche sur le standard de neutralité par rapport au temps et sur les implications théoriques potentiellement dangereuses de cette approche. La quatrième et dernière partie aborde la question de la rationalité de l'égoïsme et la possibilité ouverte par le réductionnisme de justifier rationnellement l'altruisme. Cette dernière section nous permet de nous interroger sur la proximité intellectuelle entre la philosophie de Parfit et l'utilitarisme.

Mots-clés : Parfit (Derek), identité personnelle, rationalité, utilitarisme

## **Parfit, Rational Egoism and the Issue of Personal Identity**

The present article gives an illustration of the importance of economic philosophy by showing how the concept of individual rationality depends on one's approach to personal identity. Considering the importance of the topic and the already growing literature on the subject, I decided to focus on Parfit's groundbreaking *Reasons and Persons*, published in 1984, which still remains unavailable in the French language. Parfit's reductionist approach to identity, partly inspired by Locke, radically altered conceptions of rationality. More specifically, it contributed to the questioning of the theoretical foundations of rational egoism. The first part of this study introduces and presents reductionism as an alternative to the dualist conceptions of personal identity. The second and third parts respectively discuss the impact of this view on our relation to time, and addresses the question of (im)prudence. The fourth part of this paper considers the question of whether it is necessarily rational to be egoistic. This final section also details a the way in which Parfit's philosophy can be used to support utilitarianism.

Keywords: Parfit (Derek), personal identity, rationality, utilitarianism

**JEL: A11, A12, D64, D90**

---

\*CNRS, Triangle UMR 5206, Ecole Normale Supérieure de Lyon.  
christophe.salvat@ens-lyon.fr

La question de l'identité personnelle peut, au premier abord, sembler étrangère aux préoccupations des économistes. Elle lui est pourtant centrale, et de plus en plus d'économistes en ont pris conscience au point d'être devenue un thème central de ce que l'on appelle aujourd'hui la philosophie économique<sup>1</sup>. La question de l'identité personnelle est l'une des plus discutées en philosophie économique (avec celle de la rationalité qui lui est liée). Elle répond à un besoin des économistes, celui d'aller plus loin dans leur façon de rendre-compte du choix individuel et des interactions sociales.

Derrière la question de l'identité personnelle se profile le principe de l'égoïsme rationnel, ou plus exactement celui de la rationalité individuelle, sur lequel repose toute la théorie de la décision. S'il est démontré, par exemple, que l'identité personnelle se réduit à un certain nombre de liaisons ou de connexions psychologiques, qui s'atténuent avec le temps, un agent économique devrait logiquement préférer son présent à son lointain futur, quitte à ne pas maximiser son utilité totale sur le long terme. On peut même aller plus loin. Si l'on peut montrer, comme le fait Parfit, que l'identité personnelle n'est pas importante en soi, mais que seule compte la connexité psychologique, alors ce même agent économique peut être logiquement amené à donner la préférence à un tiers plutôt qu'à son futur soi. Un examen approfondi des fondements de l'identité personnelle peut donc remettre en cause deux hypothèses fondamentales de la science économique, celle de l'intérêt personnel et celle de la neutralité par rapport au temps.

Le présent article se propose donc d'étudier l'impact de l'approche réductionniste de l'identité personnelle sur notre conception de la rationalité. Est-il possible, comme le suggère Parfit, que notre conception de la rationalité individuelle soit implicitement fondée sur une approche datée et erronée de l'identité, celle du dualisme cartésien ? Si tel est le cas, ne sommes-nous pas dans l'obligation d'interroger à nouveau frais ce qu'il nous semble être une évidence, à savoir le principe de l'égoïsme rationnel ? L'égoïsme rationnel est une formule que l'on doit à Sidgwick, et qui désigne l'ensemble des théories qui justifient les actions intéressées ('self-interested actions') et qui permettent une maximisation de l'utilité sur le long terme. En d'autres termes, l'égoïsme rationnel regroupe l'ensemble des théories qui considèrent qu'un agent est rationnel si, et seulement si, il se donne la préférence sur tous les autres individus mais qu'il ne préfère aucune période de sa vie aux autres. Ce dernier principe est connu sous le nom de neutralité par rapport au temps. Bien qu'il soit encore parfois défendu par un certain nombre de philosophes (Brink, 2011, Greene and Sullivan, 2015), la plupart des économistes considèrent une actualisation à

<sup>1</sup> Voir notamment les ouvrages de John Davis (Davis, 2003, 2010), ainsi que le numéro spécial de la *Revue de Philosophie Économique*, 'Le concept d'identité', dirigé par Alan Kirman and Miriam Teschl publié en 2004.

taux constant comme rationnelle. Cela ne fait, cependant, que peu de différence pour nous, dans la mesure où le réductionnisme permet de justifier une pure préférence pour le présent, que ni les uns ni les autres ne peuvent accepter. L'égoïsme, ou le fait de se préférer aux autres, a souvent été critiqué comme norme de rationalité, notamment par les utilitaristes qui préféreraient lui substituer un principe de neutralité par rapport à l'agence, ou selon la terminologie de Sidgwick, une bienveillance rationnelle. Mais, à l'instar de ce dernier qui fut forcé de reconnaître l'existence d'un dualisme de la raison pratique, la rationalité de l'égoïsme n'a jamais pu être véritablement contestée. Le réductionnisme offre, cependant, de nouveaux arguments qui pourraient s'avérer décisifs.

Cet article est divisé en quatre sections. Dans une première section, je présente l'approche réductionniste de l'identité telle que la défend Parfit. Je montre, notamment, que chez Parfit, ce n'est pas l'identité personnelle qui importe mais la Relation R. La deuxième section introduit la question de la rationalité, plus exactement celle de la prudence, c'est-à-dire de la neutralité par rapport au temps. Je montre qu'il est possible, selon Parfit, de ne pas respecter les règles de rationalité imposées par le principe de neutralité par rapport au temps sans pour autant être complètement irrationnel. Cela ne permet pas, cependant, d'éviter tous les comportements imprudents (section 3). La quatrième et dernière section aborde la question de l'égoïsme en tant que telle, et tente d'évaluer l'impact du réductionnisme sur la défense du principe de neutralité par rapport à l'agence. Celle-ci me permet également de montrer en quoi le réductionnisme cautionne la philosophie utilitariste à laquelle Parfit est souvent associé.

## 1. L'approche réductionniste de Parfit

Bien que peu connu de ce côté de la Manche, Derek Parfit est certainement l'un des plus importants philosophes de langue anglaise du XX<sup>e</sup> siècle. Publié il y a maintenant un peu plus de trente ans, son ouvrage *Reasons and Persons* a complètement changé notre manière de concevoir les liens entre rationalité, moralité et identité. Sa contribution à l'éthique peut se comparer à celle de Sidgwick un siècle plus tôt, ou à celle de Rawls en philosophie politique. Parfit est certes moins populaire que Rawls. Son ouvrage est, tout d'abord, incomparablement plus difficile à lire, ce qui explique sans doute l'absence de traduction française encore aujourd'hui. Ses thèses sont, ensuite, moins consensuelles et moins réconfortantes que celles de Rawls. La lecture de Parfit est souvent frustrante, car elle nous met face à l'infinie complexité des questions traitées et aux limites de la philosophie.

L'objectif de Parfit dans *Reasons and Persons*<sup>2</sup> (Parfit, 1984) est, cependant, clair : il s'agit de remettre en question le dogme de l'égoïsme rationnel que Parfit appelle les théories S (pour *Self-Interest*). Pour cela, Parfit s'attaque d'abord au postulat de la neutralité par rapport au temps (Partie II de *Reasons and Persons*). Bien qu'un certain nombre de ses arguments soient convaincants, aucun ne l'est suffisamment pour définitivement remettre en question S. La raison objective seule ne permet pas de contester le dogme de l'égoïsme rationnel. L'introduction de la question de l'identité personnelle, en revanche, peut être décisive.

La troisième partie de *Reasons and Persons*, consacrée à l'identité personnelle, est la plus longue, la plus innovante et probablement la plus discutée de son livre (avec la 'Conclusion Répugnante' de la quatrième partie). Son originalité doit beaucoup à la force logique des arguments utilisés pour défendre la thèse pourtant très contre-intuitive du réductionnisme, et aux réponses qu'il apporte à ses célèbres opposants, notamment Richard Swinburne et Bernard Williams. Mais, en dehors de son strict intérêt philosophique, l'argument non-réductionniste est important car décisif dans sa tentative de démolition du paradigme de l'égoïsme rationnel et sa défense de l'utilitarisme.

Parfit oppose deux, et seulement deux, façons de concevoir l'identité. Soit on est réductionniste, soit on ne l'est pas. En termes plus familiers, soit on accepte le dualisme, soit on le rejette. En réalité, les frontières entre ces différentes approches sont parfois plus difficiles à déterminer. Il existe, notamment, différentes approches qui ont pour seul point commun d'être opposées au dualisme. La thèse dualiste de l'identité personnelle est celle notoirement défendue par Descartes bien sûr, mais plus récemment par Richard Swinburne (Shoemaker and Swinburne, 1984). Elle stipule que le sujet ne serait se réduire à son corps mais à son essence (son esprit, son âme etc) qui n'est pas matérielle. Parfit l'appelle alternativement approche simple ou non-réductionniste. Par opposition, les approches réductionnistes considèrent que l'identité peut être déterminée en termes de continuité physiques et/ou psychologiques. John Locke a beaucoup contribué à répandre l'idée que seule comptait la continuité psychologique. Ainsi, le fait que je puisse me souvenir de mes expériences passées prouve que l'individu qui a vécu ces expériences et celui qui se les remémore est le même. Mais cette théorie se confronte très vite à ses propres limites. Le fait que je ne me souvienne plus de certaines expériences passées ne signifie pas nécessairement que je ne suis pas celui qui les a vécues. Pour répondre à ces problèmes, la plupart des philosophes ont ajouté le critère de continuité physique à celui de continuité psychologique. Il est évident qu'il n'y a pas de continuité psy-

<sup>2</sup> Toutes les références à cet ouvrage renvoient à l'édition *paperback* de 1987.

chologique possible sans continuité physique, celle de notre cerveau à tout le moins. Les philosophes Bernard Williams et Thomas Nagel font ainsi de la conservation du cerveau un élément majeur de l'identité personnelle (Williams, 1976, Nagel, 2012). Mais la compatibilité de ces deux principes peut aussi s'avérer difficile ainsi que le montre très bien Parfit dans ses différents exemples imaginaires (dans quelle mesure, par exemple, un individu A auquel on aurait greffé le cerveau de B, et qui se souviendrait de toutes les expériences passées de celui-ci, peut-il devenir B?). Cette difficulté théorique a conduit certains penseurs à totalement rejeter le principe de continuité psychologique pour ne garder que le principe de continuité physique (Olson, 1997).

La simple continuité psychologique ne suffit pas à établir l'identité des personnes dans le temps. Celle-ci doit également reposer, nous dit Parfit, sur un degré minimum de *connexité psychologique* ('*psychological connectedness*'). La connexité psychologique peut être définie comme la relation directe, plus ou moins forte, qu'il existe entre deux états psychologiques. A la différence de la simple continuité, la connexité psychologique admet des degrés, et donne ainsi la profondeur qui manquait chez Locke à la notion de continuité. La continuité psychologique peut alors se définir, quant à elle, « comme une série de fortes connexités psychologiques, qui se chevauchent dans le temps (§78, 206) ».

L'intérêt de la définition de Parfit est qu'il est dorénavant possible d'approcher la notion d'identité personnelle d'un point de vue qualitatif. L'existence de quelques souvenirs qui se recoupent dans le temps ne suffit pas à établir une relation d'identité. En revanche, l'existence de fortes proximités psychologiques, qui se succèdent et se recoupent dans le temps, permet d'envisager une telle relation. Naturellement, et j'aurais l'occasion d'y revenir ci-dessous, cette relation d'identité est d'une nature différente de celle initialement conçue. Chez Locke, comme chez la plupart des réductionnistes (et *a fortiori* des non-réductionnistes), l'identité est déterminée par l'existence d'un lien psychologique continu. Dans sa version stricte, ce lien est direct mais on a vu qu'il pouvait également être indirect. Dès lors que l'existence de ce lien est établie, l'identité dans le temps est assurée. L'identité personnelle est donc '*tout ou rien*' ('*all-or-nothing*'). La notion de connexité, en revanche, repose sur une évaluation qualitative du lien psychologique (qui, lui, est toujours direct). Pour Parfit, on ne peut parler de continuité psychologique que s'il existe une série ou une chaîne de fortes connexités, qui se chevauchent dans le temps. Rien, cependant, ne nous permet de déterminer avec précision une forte d'une faible connexité. Lorsque les connexions directes entre deux états psychologiques sont *suffisamment* nombreuses, on peut parler – nous dit Parfit – de forte connexité (§78, 206).

Le fait qu'il existe entre moi et une autre entité une forte continuité et connexité psychologique, ce que Parfit appelle une *Relation R*, est plus important que le fait de savoir s'il s'agit encore de moi ou non. Parfit peut donc affirmer que l'identité personnelle n'est pas ce qui importe. Ce qui compte vraiment est la Relation R, c'est-à-dire la connexité et/ou continuité psychologique, Relation qui peut soit être le fait d'une cause normale (dans le cas d'une vie ordinaire) soit le fait de n'importe quel type de cause (dans le cas de la télétransportation ou de la greffe de cerveau par exemple). Dans la plupart des cas, les cas normaux, la Relation R recouvre la notion d'identité personnelle. Parfit s'appuie donc, dans un premier temps, sur le cas de R avec le type normal de cause ('with the right kind of cause') (§79, 215). Il la réintègre ensuite dans R comme étant simplement un cas parmi d'autres. On peut ainsi avoir l'impression que la question de l'identité est fondamentale. Elle ne l'est, dans ce cas, que par coïncidence.

Nous avons tendance à penser qu'il ne peut y avoir qu'une seule façon de répondre à la question 'Est-ce que je vais mourir?', soit par 'oui' soit par 'non'. De la même manière, nous avons tendance à penser que notre identité est sans équivoque. Ces réactions révèlent notre non-réductionnisme instinctif. Le réductionnisme nous enseigne, nous dit Parfit, que l'identité personnelle n'est pas déterminée et que, par conséquent, la question de notre survie est moins catégorique qu'il n'y paraît<sup>3</sup>. Puisque la thèse du réductionnisme semble acquise, alors il semble logique pour Parfit que nous devions abandonner nos certitudes identitaires, et accepter le fait qu'on ne puisse pas répondre à des questions telles que 'Est-ce que je vais mourir?'. Cela est, cependant, aller un peu vite en besogne et certaines résistances se manifestent au sein même du réductionnisme. Parmi celles-ci, celles du philosophe britannique Bernard Williams ont particulièrement attiré l'attention<sup>4</sup>.

Le caractère très contre-intuitif de cette thèse peut expliquer les nombreuses résistances qu'elle rencontre<sup>5</sup>. Même après un change-

<sup>3</sup> Ceci est contesté par Shoemaker qui affirme que le critère de continuité (et connexité) psychologique n'est pas théoriquement incompatible avec la thèse du non-réductionnisme (Shoemaker, 1997, 13).

<sup>4</sup> Pour Williams, la continuité corporelle prime sur la simple continuité psychologique. Dans son article 'The Self and the Future' (1970), Williams avait critiqué le raisonnement (ou plus exactement l'expérience de pensée) selon lequel deux individus A et B pouvaient échanger leurs identités en échangeant simultanément les données de leur cerveau dans une machine prévue à cet effet (Williams, 1970). Dans son article 'Experiences, subjects, and conceptual schemes' (1999), Parfit affirme que Williams est 'hyper-réductionniste', ce qui fait de lui un dissident du réductionnisme (Parfit, 1999). L'article, traduit par Sophia Mo, est disponible en français sur la page web <http://semahp.blogspot.fr/2014/10/traduction-de-derek-parfit-experiences.html>.

<sup>5</sup> Thomas Nagel aurait ainsi affirmé que même si la thèse du réductionnisme était vraie, il serait psychologiquement impossible d'y croire (§94, 274).

ment physique et/ou psychologique significatif, il nous paraît certain qu'on puisse toujours répondre à la question 'Est-ce toujours la même personne ?' (et, dans la plupart des cas, la réponse sera 'oui'). Ce n'est pourtant pas systématiquement le cas. Nous acceptons ainsi très bien le principe d'indétermination dans le cas d'une nation ou d'un club (§79, 213). Les membres d'un club comme d'une nation changent, leurs règles de fonctionnement changent aussi, parfois ils changent même de nom. Peut-on dire, par exemple, que le club de l'Athenaeum à Londres, l'un des plus sélectifs d'Angleterre, est le même aujourd'hui – après qu'il ait ouvert ses portes à la gente féminine – qu'à sa création en 1824 ? Beaucoup de ses anciens membres en douteraient (et peut-être certains de ses présents membres en doutent-ils encore). La France est-elle la même sous François Hollande que sous Napoléon 1<sup>er</sup>, qui lui a donné, entre autres le code civil et l'ensemble de sa structure administrative ? Il n'y a pas plus de sens à répondre à ces questions qu'à celles de l'identité personnelle. Nous l'admettons (généralement) dans un cas, mais pas dans l'autre.

La thèse de l'indétermination de l'identité a deux implications théoriques importantes. Tout d'abord, elle implique – ou peut impliquer – que les relations psychologiques et physiques qui forment ce que l'on désigne généralement par le terme d'identité peuvent en fait être décrites sans référence à la personne. « Bien que les personnes existent », écrit Parfit, « il est possible de donner une description *complète* de la réalité *sans* prétendre que les personnes existent » (§79, 212). Cette affirmation a suscité beaucoup de commentaires et de confusions, obligeant Parfit à préciser qu'il ne défendait en aucun cas une version impersonnelle de l'identité<sup>6</sup>. Je ne traiterai pas de cette question ici. La deuxième conséquence théorique du réductionnisme (ou de l'indétermination puisque l'une ne va pas sans l'autre chez Parfit) est que l'identité personnelle n'a que très peu d'importance morale et/ou rationnelle. La question de l'importance morale de l'identité personnelle renvoie notamment aux débats portants sur la séparabilité des personnes et la justice distributive. Elle ne sera pas traitée ici non plus. Celle de l'importance rationnelle de l'identité porte sur ce qu'on appelle communément la prudence, c'est-à-dire le rapport que l'on entretient vis-à-vis de notre futur bien-être. C'est l'un des fondements du dogme de l'égoïsme rationnel, qui considère que la rationalité individuelle consiste à se préférer aux autres mais à ne pas faire de préférence entre son bien-être présent et son bien-être futur. Mais, le principe prudentiel ne peut se justifier qu'à condition que je sois la même personne aujourd'hui et dans le futur (sinon je viole le premier postulat puisque je ne me préfère pas aux autres). Que

---

<sup>6</sup> Dans le sens où les expériences n'ont pas de sujets, et les pensées n'ont pas de penseurs.

devient donc ce principe si l'on accepte la thèse de l'indétermination de l'identité ? C'est l'objet de la prochaine section.

## 2. Identité personnelle et rationalité

La section précédente a montré que, pour Parfit, le problème de l'identité personnelle est une question qui, au mieux n'a pas de réponse, au pire qui n'a pas de sens. Elle ne peut donc justifier ou motiver nos actions dans le temps. Adopter une approche réductionniste de l'identité peut donc se révéler lourd de conséquence sur notre conception de la rationalité. Si j'admets, après avoir lu Parfit par exemple, que ce qui importe ce n'est pas l'identité dans le temps mais la continuité et/ou connexité psychologique, avec n'importe quel type de cause, (Relation R), alors je pourrais être justifié (1) de préférer mon présent à mon futur et (2) de ne plus me préférer aux autres. Autrement dit, le réductionnisme (tel que l'entend Parfit) remet à la fois en cause les deux piliers de l'égoïsme rationnel que sont la neutralité par rapport au temps et la non-neutralité par rapport à l'agence.

Considérons, tout d'abord, le cas de la neutralité par rapport au temps. Dans quelle mesure la perspective de nos plaisirs et de nos peines peut-elle influencer nos décisions si nous ne nous les percevons plus comme *nos* futures peines et plaisirs ? Pour un non-réductionniste, comme Swinburne, il est évident que la seule raison pour laquelle nous nous soucions de notre futur, est précisément qu'il s'agit de *notre* futur. Si, maintenant, on accepte la thèse réductionniste de Parfit, il ne s'agit plus à proprement parler de *notre* futur mais du futur d'une personne avec laquelle nous entretenons une continuité (et/ou connexité) psychologique plus ou moins grande. Il est alors possible d'affirmer que, si l'on adopte une position réductionniste, alors on a plus de raison de se soucier de son futur. C'est ce que Parfit appelle la *Proposition Extrême* ('*The Extreme Claim*'), et dont il attribue la première formulation à Butler (§102, 307).

Il est important de noter que Parfit avait initialement rejeté cette proposition. Dans son article 'Personal Identity and Rationality' (1982), Parfit avait tout d'abord soutenu l'idée que les non-réductionnistes avaient tort de croire que le fondement du souci de soi était dans ce quelque chose de plus qui constituait selon eux l'identité personnelle. Dans une métaphore très inspirée, il comparait alors cette 'chose profonde' ['deep further fact'] au Soleil qui éblouit et cache la présence constante, mais du coup moins remarquée, de la Lune, censée représenter la continuité psychologique (Parfit, 1982, 230). Sensible aux critiques de Wachsberg (Wachsberg, 1983), Parfit revient sur sa position dans *Reasons and Persons*, et admet que d'un point de vue non-réductionniste, il est légitime de penser – qu'en devenant réductionniste – nous perdrons toute raison de nous pré-



occuper de notre futur<sup>7</sup>. D'un autre côté, les non-réductionnistes peuvent arguer du fait qu'on se préoccupe tout aussi bien, et parfois mieux, du futur des personnes qui nous sont chères que de notre propre personne, et que ce n'est donc pas l'identité personnelle mais la Relation R qui compte dans notre relation au futur. C'est ce que Parfit appelle la *Proposition Modérée* ('*The Moderate Claim*'). L'argument est un peu troublant. Si, en se convertissant au réductionnisme, un non-réductionniste pense devoir perdre toute préoccupation pour son futur, c'est qu'il continue à penser comme un non-réductionniste.

Selon la *Proposition Modérée*, la justification de se soucier de son futur bien-être dépend de la connexion/connexité psychologique que l'on entretient avec son futur 'soi'. Elle peut donc valider comme invalider le principe de l'égoïsme. La Proposition Extrême, en revanche, l'invaliderait absolument. De ce point de vue, elle constituerait l'argument ultime contre S. C'est probablement la raison pour laquelle il ne la rejette pas immédiatement, même si cela signifie faire appel à un autre cadre théorique que le sien pour défendre sa thèse, ce qui peut sembler étrange.

Les partisans de S sont favorables au principe de neutralité par rapport au temps. Nous avons déjà eu l'occasion de noter que derrière cette formulation un peu abstraite, ils voulaient en fait signifier qu'une personne rationnelle devrait également se soucier de toutes les parties de son futur. C'est ce que Parfit appelle la *Règle de l'Importance Egale* ('*The Requirement of Equal Concern*') (§103, 313). Ce principe est, en quelque sorte, une version simplifiée de la neutralité par rapport au temps à laquelle on mettrait à part les faits ayant eu lieu dans le passé et ceux ayant lieu dans le présent. Ontologiquement, cette règle repose sur un principe évident (si l'on n'est pas réductionniste) à savoir que « toutes les parties du futur d'une personne sont également des parties de son futur » (§104, 315). C'est ce que Parfit appelle le *Truisme* ('*The Truism*').

Le truisme bénéficie d'un certain attrait intuitif mais il repose sur un faux-semblant logique : la continuité n'implique pas l'égalité. Prenons l'exemple de la parenté : si l'on remonte très loin dans le temps tous les hommes, et même tous les mammifères, sont également mes parents : on pourrait théoriquement établir une continuité biologique. Il serait cependant absurde d'en conclure qu'ils sont mes parents au même degré, ou de façon égale ! De la même façon, ce n'est pas parce que toutes les parties de mon futur sont également des parties de mon futur qu'elles comptent également pour moi.

Nous avons vu dans la section précédente que la Relation R s'appuyait sur deux types de relation psychologique, la connexion et

<sup>7</sup> L'argument est un peu troublant. Si, en se convertissant au réductionnisme, un non-réductionniste pense devoir perdre toute préoccupation pour son futur, c'est qu'il continue à penser comme un non-réductionniste.

la connexité. La première est une relation transitive qui établit que je suis lié à telle ou telle partie de mon futur en raison de toutes les parties intermédiaires de mon futur avec lesquelles je suis déjà lié. Ainsi dans la séquence  $i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8, i_9, i_{10}$ , je peux dire que la personne  $i$  en temps 1 est liée à la personne  $i$  en temps 10 parce que cette dernière est psychologiquement proche de  $i_9$ , qui est elle-même est proche de  $i_8$  ... jusqu'à  $i_1$ . C'est l'identité lockéenne. L'identité se fonde sur la mémoire : je sais que  $i_{10}$  est la même personne que  $i_1$  parce que  $i_{10}$  se souvient d'avoir été  $i_9$ , qui se souvient d'avoir été  $i_8$  etc. C'est, nous l'avons vu, ce que Parfit appelle la continuité psychologique. Dans la mesure où chaque période (1, 2, 3... 10) tient une place également importante dans la séquence qui fait l'identité de  $i$ , le truisme est vrai : pour  $i_1$  toutes les parties de son futures sont également les parties de son futur. En ce qui concerne son identité, aucune ne joue un rôle plus important que l'autre.

Mais on a tort, nous dit Parfit, de nous contenter de cette interprétation de l'identité. Il est également possible de considérer séparément les éléments de cette séquence. C'est ce que Parfit appelle la connexité psychologique. Ainsi je peux m'interroger sur la relation entre  $i_1$  et  $i_6$  et entre  $i_3$  et  $i_{10}$  indépendamment l'une de l'autre. Il n'y a plus nécessairement de transitivité entre mes mois temporels, et ce n'est plus la même chose de se demander s'il y a identité entre  $i_1$  et  $i_{10}$ , d'une part, et  $i_3$  et  $i_6$ , d'autre part. Ce qui compte maintenant c'est la relation *directe* qu'entretient  $i_1$  avec  $i_6$  ou  $i_3$  avec  $i_{10}$ . Or certains éléments de cette relation tiennent à différents degrés dans le temps. Bien que les séquences ( $i_1$ - $i_{10}$ ) et ( $i_3$ - $i_6$ ) soient *également* continues, il n'y a pas nécessairement le *même* degré de connexité entre  $i_1$  et  $i_{10}$  qu'entre  $i_3$  et  $i_6$ . Ce n'est pas parce que toutes les parties de mon futur sont également des parties de mon futur que je dois les traiter également. Autrement dit, la continuité psychologique justifie le fait qu'un individu se préoccupe de son futur. La connexité, quant à elle, justifie le fait qu'il s'en préoccupe moins. Il est donc maintenant possible de rejeter la *Règle de l'Égale Importance*, et d'affirmer qu'il est rationnel pour une personne de moins se préoccuper de son futur lointain que de son futur proche dans la mesure où (comme c'est généralement le cas) elle est plus étroitement connectée avec son futur proche qu'avec son futur lointain.

Cela ne constitue pas, bien sûr, une réfutation de S, mais cela conduit néanmoins à en réviser significativement les prémisses. Un pan du principe de neutralité par rapport au temps s'effondre. Il est vrai qu'en pratique la plupart des philosophes et des économistes admettent qu'il est rationnel de préférer le présent au futur et le futur proche au futur lointain. L'actualisation du temps, lorsqu'elle se fait à taux constant, n'est pas contraire au modèle de rationalité standard, même si elle constitue une légère entorse à la théorie classique de la prudence. Mais, et c'est un point important, l'approche réductionniste

de Parfit conduit à actualiser les degrés de connexité psychologique plutôt que le passage du temps. Si ma connexité psychologique s'effrite à taux constant, cela ne fait pas de différence. Mais, il n'y a aucune raison que ce soit le cas. On peut également imaginer des points de retournements (ce qui est naturellement impossible dans le cas du passage du temps). Pour reprendre l'exemple du jeune aristocrate Russe<sup>8</sup>, il n'est pas impossible qu'il soit politiquement révolutionnaire dans sa jeunesse, conservateur devenu adulte et que, désenchanté, il redevienne progressiste dans ses vieux jours. La Relation R sera alors plus étroite sur le long terme que dans le moyen terme. De la même façon, il n'est pas impossible qu'il existe un degré de connexité psychologique plus important entre  $i_1$  et  $i_{10}$  qu'entre  $i_3$  et  $i_6$ . Les deux formes d'actualisation ne sont donc pas comparables.

Cela a deux implications majeures. La première c'est que le biais vers le proche est irrationnel s'il est le produit d'une actualisation du passage du temps, mais qu'il ne l'est pas nécessairement s'il est le produit d'une actualisation du degré de connexité psychologique. Et cela reste vrai même s'il n'y a pas de point de retournement. Je m'explique. Le passage du temps est constant. Pour qu'il y ait un renversement des préférences de l'individu, il faut donc qu'il y ait une variation du taux d'actualisation. Le degré de connexité psychologique, lui, n'est pas constant, ou plus exactement il n'a aucune raison de l'être. Le degré de connexité du jeune aristocrate Russe peut ainsi rester très élevé entre 20 et 25 ans puis s'effondrer à mesure qu'il approche de ses 30 ans. Plus il approche de ses 30 ans, plus la Relation R se distend. Une forte connexité psychologique fera l'objet d'une faible actualisation et inversement. Si je suppose que ma connexité s'effrite dans le temps à un taux croissant, alors je l'actualiserais également de façon croissante, et mon taux d'actualisation sera, lui, décroissant. Le temps ne se distend pas et nous le savons. Si nous actualisons en fonction de notre degré de connexité, et non simplement en fonction du temps, comme on le suppose généralement, nous avons des raisons de faire des choix dynamiquement incohérents. L'une des implications majeures du réductionnisme est donc de montrer que les individus n'agissent pas nécessairement dans leur meilleur intérêt et, qui plus est, qu'ils ont de bonnes raisons d'agir ainsi.

---

<sup>8</sup> L'exemple de Parfit est le suivant : au dix-neuvième siècle, un jeune aristocrate est destiné à hériter un vaste domaine terrien. Ayant des idéaux socialistes, il souhaite – maintenant – faire don de ces terres aux paysans qui les cultivent. Il sait, cependant, qu'en vieillissant, ses allégeances politiques changeront. Il signe donc un document le forçant à donner ses terres le jour où il héritera (§110, 327).

### 3. La question des 'sérieuses imprudences'

Si le réductionnisme permet d'attaquer S, et indirectement la rationalité de l'égoïsme, il peut également apparaître comme une justification de l'imprudence, et ainsi donner des arguments aux partisans de la *Proposition Extrême*. Or, c'est une direction dans laquelle Parfit ne veut pas aller. La théorie de la rationalité qu'il défend, CP, ne peut se permettre de justifier ce que Parfit appelle les 'sérieuses imprudences'. Une théorie qui motiverait toute sorte d'actions propres à se mettre en danger ne peut décemment se présenter comme un modèle de rationalité. C'est la raison pour laquelle, et bien que Parfit ait frontalement attaqué S sur ce point, CP est neutre par rapport au temps. Autrement dit, pour S comme pour CP, nous n'avons aucune raison de moins nous préoccuper de notre futur lointain simplement parce qu'il est lointain. Nous avons, en revanche, dans le cas de CP, certaines raisons de moins nous en préoccuper si nous ne sommes que superficiellement connecté avec celui que nous serons dans ce futur lointain. Mais, à degré de connexité égal, il serait absurde de préférer une période de notre vie à l'autre. Reste que CP pourrait justifier un certain nombre d'actions dites imprudentes. Pour corriger cette tendance, Parfit envisage de compléter CP par un volet moral et un volet évolutionniste. Commençons par ce dernier.

Pour un réductionniste, l'identité personnelle n'est pas une question importante. Cela ne signifie pas que les personnes n'existent pas, mais qu'il est vain de rechercher à savoir si on reste bien la même personne avec le temps qui passe. Tout ce qu'on peut dire avec certitude, c'est qu'un individu entretient des continuités physiques et psychologiques plus ou moins étroites avec les personnes qu'il a été dans le passé ou qu'il sera dans le futur. L'identité n'admet pas toujours de réponses définitives (telles que 'oui, c'est toujours la même personne' ou 'non, celui qu'il a été a cessé d'exister'), mais doit le plus souvent s'apprécier en termes de degrés. Mais la mort, pourrait-on dire, n'est pas une question de degré. Je sais que lorsque le moment viendra je n'existerai plus, et cela doit nécessairement avoir une influence sur mes croyances et mes actions présentes. Et si la question de ma non-existence est importante pour moi, alors nécessairement la continuation de mon existence doit l'être tout autant. A cela Parfit répond que « [b]ien qu'il y aura [après ma mort] beaucoup d'expériences, aucune de ces expériences ne sera directement connectée à mes expériences présentes à travers des suites de connections directes telles que celles qu'impliquent la mémoire-expérience, ou l'accomplissement d'intentions passées. Quelques-unes de ces expériences futures peuvent être connectées à mes présentes expériences d'une façon moins directe. Il y aura plus tard quelques souvenirs de ma vie. Et peut-être y aura-t-il quelques pensées qui auront été influencées par les miennes, ou des choses réalisées grâce à mes conseils. Des relations

plus directes entre mes expériences présentes et les expériences futures seront rompues par ma mort, mais un certain nombre d'autres relations ne le seront pas. La seule chose est qu'il n'y aura pas de personne vivante qui soit moi. Dès lors que j'ai compris cela, ma mort me semble moins terrible. » (§95, 281).

Nous n'avons peur de la mort que parce que nous voyons en elle la fin de ce que nous sommes. Mais le réductionnisme nous dit que nous ne sommes qu'une série de continuités physiques et psychologiques. Or celles-ci, ou plutôt un certain nombre d'entre elles, continuent à exister à travers d'autres personnes vivantes. Il n'y a, après tout, entre l'exemple de la télétransportation et la réalité qu'une question de degré. Lorsque je suis télétransporté, un être physiquement différent – bien que similaire – de moi se souvient de mes expériences passées, partage mes valeurs et mes goûts et s'engage à accomplir ce que moi-même je désirais accomplir. Si je suis réductionniste, je ne perçois pas cette destruction de mon corps comme équivalente à ma mort<sup>9</sup>. Je peux ne pas la préférer à ma survie ordinaire, mais je la préférerais certainement à ma mort ordinaire. Imaginez maintenant que je sois proche de la mort, c'est-à-dire de la destruction physique de mon corps, mais qu'il existe une jeune personne qui m'est chère, à qui j'ai raconté ma vie dans ses moindres détails, à qui j'ai confié tous mes sentiments, à qui j'ai inculqué mes valeurs et ma philosophie de la vie, à qui je lègue maintenant tous mes biens, et que je sache que cette personne va continuer à habiter et entretenir ma maison, et surtout à veiller et à entretenir mon jardin, qui constitue ma plus grande fierté et dans lequel mes cendres seront dispersées. N'est-ce pas là une raison de moins craindre la mort ? A part son corps, qui est dissemblable du mien, cette personne est-elle vraiment différente de la personne que je serais censée être après être télétransporté ?

Non, en tout cas, pas de façon radicale. Ce qui compte, comme on a pu le voir dans la section précédente, c'est la persistance d'une continuité et/ou d'une connexité psychologique à travers le temps, avec n'importe quel type de cause, ce qu'il appelle la Relation R. Jusqu'à présent, nous n'avons pas prêté attention à l'addendum 'avec n'importe quel type de cause'. Il est temps de s'y arrêter ici. Si je considère que ce qui importe c'est la Relation R, avec une cause normale, ma continuité/connexité psychologique implique nécessairement une continuité physique, au moins celle du cerveau. Du point de vue de ma survie, cela ne change donc rien. Toute continuité psychologique

---

<sup>9</sup> Parfit conteste l'idée que l'on puisse être réductionniste et regarder la continuité physique comme primordiale (§96, 284). Cela explique notamment sa réticence à compter Williams parmi les réductionnistes (§91, 268). Nombreux sont ceux, pourtant, qui se réclament du réductionnisme et qui voient dans la continuité physique du corps (Williams, Jarvis Thompson) ou du cerveau (Nagel), le fondement de l'identité personnelle.

sera impossible après ma mort. Et l'expérience de la télétransportation serait, de ce point de vue, équivalente à la mort. Si je considère maintenant que la Relation R me permet, dans une certaine mesure, de continuer à exister bien que non nécessairement dans mon corps et mon cerveau d'origine, alors il importe peu que ce corps soit identique ou non à celui que j'avais. En d'autres mots, si j'accepte le fait que mes souvenirs, mes pensées et mes intentions soient télétransportés, dans un corps et un cerveau qui n'est pas le mien, je devrais également regarder avec sympathie l'idée que quelqu'un d'autre puisse continuer à se souvenir de mes expériences passées, partager mes valeurs et se proposer d'accomplir mes projets. C'est ce que Parfit appelle la Relation R *avec n'importe quel type de cause*.

L'implication pratique de cette thèse est énorme. Cela change, ou peut potentiellement changer, notre rapport à la mort. Si je considère que mon clonage est aussi bon que la survie ordinaire (voire meilleur puisque le nombre de mes années de vie est démultiplié), alors je peux considérer que l'existence d'une continuité/connexité psychologique entre mon moi présent et d'autres personnes constitue – à un degré moindre – une forme de survie extracorporelle. Ce n'est probablement pas équivalent au fait de rester en vie, mais c'est tout de même moins effrayant que de ne plus exister du tout<sup>10</sup>.

Disons plutôt que cela pourrait théoriquement changer notre rapport à la mort, mais qu'en pratique cela ne le change pas. La première raison est, ainsi que Parfit le souligne à plusieurs reprises, que les hommes sont, dans leur très grande majorité, instinctivement non-réductionnistes. Et l'on pourrait ajouter que, parmi ceux qui sont réductionnistes, une minorité partage la position radicale de Parfit ! Mais, serait-ce le cas que cela ne changerait probablement rien, ou presque rien, à notre rapport avec la mort. Nous avons instinctivement peur de la mort, si ce n'était pas le cas l'espèce humaine serait d'ailleurs depuis longtemps éteinte (§102, 308). Nous pouvons nous convaincre que cette peur n'est pas rationnelle, qu'elle n'est pas théoriquement justifiée, elle continuera à nous effrayer. A tout le moins, notre conception de l'identité peut-elle dédramatiser la perspective de notre propre mort, mais elle ne pourra jamais nous rendre totalement indifférents à son égard. Nous sommes naturellement conditionnés à craindre notre mort, et par conséquent à éviter toutes les actions qui pourraient la précipiter. Cela constitue une première réponse possible à la Proposition Extrême.

---

<sup>10</sup> On pourrait, cependant, opposer à Parfit que ce n'est pas parce que cette vision est réconfortante qu'elle est vraie. Le fait qu'il existe quelqu'un qui se souvienne de vous et qui continue votre œuvre ou qui partage vos idéaux n'est pas équivalent au fait d'être vivant. L'important, objecte ainsi Mark Johnston, n'est pas que telle ou telle expérience soit vécue mais bien qu'elle soit vécue par *vous*. Et nul besoin, d'ailleurs, d'abandonner le réductionnisme pour rétablir le principe du souci de soi. L'un et l'autre sont parfaitement compatibles (Johnston, 1997).

Une seconde possibilité est de voir dans la prudence une norme morale plutôt qu'une norme de rationalité. Parfit suggère trois justifications morales possibles de la prudence (§106). La première est d'inspiration utilitariste, la deuxième d'inspiration kantienne et la troisième, et bien disons d'inspiration parfitienne. Du point de vue utilitariste, les imprudences sont *moralement* condamnables dans la mesure où elles amoindrissent la somme totale des utilités. Or, à moins que l'imprudence d'une personne profite aux autres, c'est presque toujours le cas. En plus les désutilités personnelles qu'elle génère, l'imprudence se caractérise par de fortes externalités négatives. L'imprudence d'un individu a généralement un coût affectif et/ou financier pour les proches mais aussi pour la collectivité qui doit souvent en payer le prix (assistance médicale et/ou sociale, primes d'assurances, surcroît de réglementation etc..). D'un point de vue kantien ou plutôt néo-kantien, je ne dois pas faire de différence entre moi et les autres. On notera le retournement ironique opéré ici par Parfit. Alors que l'argument kantien est destiné à convaincre du bien-fondé de l'altruisme et du sacrifice de soi, il est utilisé ici pour justifier des actions égoïstes. Le raisonnement est le suivant : s'il n'y a pas de différence entre moi et les autres, et que je me dois aux autres, alors je me dois également à moi-même. Certes, je ne me dois pas plus à moi-même qu'aux autres, mais suffisamment pour ne pas me mettre dans des situations difficiles<sup>11</sup>.

La troisième justification morale de la prudence est la suivante. S'il n'existe pas ou peu de connexité psychologique entre mon moi présent et mon moi futur, alors je dois me comporter envers celui-ci comme je me comporterais envers quelqu'un qui n'existe pas encore. Autrement dit, d'un point de vue moral, il n'est pas injustifié d'affirmer que les personnes-dans-le-futur ont le même statut que les générations à venir. Donc, si nous partons du principe que nous avons un certain devoir moral envers les générations à venir, alors nous avons également un devoir moral envers notre propre futur. Le problème est que Parfit ne parvient pas à démontrer, dans la quatrième et dernière partie de *Reasons and Persons*, l'existence d'un devoir moral des générations présentes envers les générations futures. Par conséquent, il ne peut pas utiliser non plus cet argument pour le devoir de prudence personnelle.

Je dirais, cependant, qu'il importe peu *in fine* de savoir quelle justification est la bonne. Le véritable problème est qu'il ait ici recours à

---

<sup>11</sup> L'argument mériterait en fait d'être approfondi car il repose implicitement sur le postulat que le paternalisme est toujours légitime. Je suis en droit d'intervenir pour mon bien-être futur car je suis en droit d'intervenir pour le bien-être futurs des autres, et je ne suis pas différent d'eux. Mais le paternalisme n'est pas justifiable, ou l'est difficilement, chez Kant. Il y a une différence entre agir comme si je n'étais pas différent des autres et agir pour moi comme j'agirais pour les autres. Or, Parfit n'élabore pas sur cette différence.

un argument moral, quel qu'il soit. Jusqu'à présent, Parfit a, en effet, été très attentif à ne pas mélanger les arguments d'ordre moral et ceux d'ordre logique, et à discuter la théorie de la rationalité *S* sur le strict plan de la rationalité. Il aurait, tout aussi bien, pu utiliser l'argument moral, qu'il considère d'ailleurs être le meilleur, dès le départ. Il s'y est refusé. Pourquoi, alors, l'utiliser à la fin, et compromettre par-là l'édifice déjà élevé ? Quant à l'argument évolutionniste, sa principale faiblesse est d'être *ad hoc*. Alors que l'argument moral tente de trouver des justifications possibles à la prudence, l'argument évolutionniste fait état d'une tendance à la prudence sans pour autant tenter de lui donner une justification. Or, et c'est ce qui – de mon point de vue – rend l'argument moral encore plus critiquable, l'argument évolutionniste suffisait à contrer la Proposition Extrême. Si, en effet, les individus luttent instinctivement pour leur survie, y compris lorsque le degré de connexité psychologique avec leur futur ne le justifie pas, alors il n'y a pas de raison que cette réaction prudentielle ne tienne pas à un degré moindre pour toutes les actions qui menaceraient la bonne conservation future de l'individu. Il est donc possible d'affirmer que le réductionnisme, bien que pouvant justifier des actions contraires, ne conduit pas nécessairement les individus à commettre de 'sérieuses imprudences'.

On peut aller plus loin, et se demander si l'erreur initiale de Parfit n'a pas été de ne pas directement intégrer l'argument évolutionniste dans la Relation *R*. S'il modifie notre rapport à la mort, comment ne peut-il pas modifier notre rapport à l'identité ? Naturellement, tout dépend du contenu normatif que Parfit entend donner à la Relation *R*. Mais *R* est-il destiné à occuper une fonction normative ? Parfit, me semble-t-il, ne dit rien à ce sujet. La finalité de la question de l'identité est de rendre compte du comportement intéressé des individus, et non pas de leur donner un argument supplémentaire (et superfétatoire) de s'intéresser à eux-mêmes. C'est ce que font les non-réductionnistes avec ce 'quelque chose en plus' qui constitue selon eux l'essence de l'identité, et c'est ce que font les réductionnistes (parfitiens) avec la Relation *R*. Il est donc possible d'intégrer l'argument évolutionniste dans *R*, bien que celui-ci n'ait aucune portée prescriptive.

Cela semble indiquer qu'il y a une faille dans l'approche réductionniste de Parfit, qui ne fait pas suffisamment place à la continuité physique. Le spectre psychologique avait pourtant démontré l'impossibilité d'exclure totalement toute caractéristique physique de notre approche de l'identité (§84). Parfit semble d'ailleurs l'admettre à plusieurs occasions, lorsqu'il fait du cerveau le vecteur de la continuité/connexité psychologique. Mais la conservation physique du cerveau, et *a fortiori* d'une partie du cerveau, est un élément insuffisant pour rendre compte du comportement naturellement prudentiel des hommes. On peut, bien sûr, associer la destruction du



cerveau à la mort ordinaire, et défendre l'idée que derrière notre peur instinctive de la mort se trouve, en fait, la peur – rationnellement justifiée – de la destruction de notre cerveau et de la rupture de la continuité psychologique. Cela pourrait d'ailleurs donner une certaine épaisseur à l'argument évolutionniste de Parfit. Mais cela ne n'éclairerait en rien la répugnance naturelle des hommes à mettre en jeu leur intégrité physique, y compris en absence de connexité psychologique. La prise en considération de l'intégrité physique dans R, en revanche, permettrait d'appréhender ce genre de comportement. Parfit a naturellement de bonnes raisons de s'y opposer. En ne faisant reposer R que sur la continuité/connexité psychologique, il peut justifier l'altruisme, ou plus exactement le fait de préférer un autrui présent à son soi futur. Intégrer des éléments physiques dans R pourrait remettre en cause ce résultat important.

#### 4. Parfit et la 'désintégration partielle' de la personne

L'approche réductionniste de l'identité personnelle de Parfit a, on l'a vu, des implications majeures sur notre rapport au temps, et peut notamment remettre en cause le standard de rationalité que constitue la neutralité par rapport au temps. Pour Parfit, son véritable enjeu est, cependant, beaucoup plus important : elle remet en cause le postulat de l'égoïsme du principe de rationalité individuelle. Le réductionnisme peut donc se révéler, sur ce point précis, un atout décisif pour la défense de l'utilitarisme<sup>12</sup>.

Pour les utilitaristes, il n'y a pas de différence entre maximiser l'utilité d'une personne et maximiser l'utilité d'un groupe de personnes. Lorsque je (à supposer que je sois utilitariste) décide de maximiser l'utilité, je ne me pose pas la question de savoir qui en profite. Supposons, avec Parfit, que nous devons imposer une épreuve (par exemple une opération chirurgicale) à un jeune enfant soit (1) pour

---

<sup>12</sup> La philosophie utilitariste, telle qu'elle est développée par Sidgwick, défend le principe de bienveillance impartiale, c'est-à-dire l'idée selon laquelle est juste toute action qui tend à améliorer le bien-être des individus, quels qu'ils soient. Elle repose sur deux postulats *a priori* incompatibles, l'égoïsme rationnel (toutes les actions sont motivées par l'intérêt personnel) et l'impartialité (personne ne compte plus que les autres, y compris soi-même). Sidgwick adopte une approche non-réductionniste de l'identité personnelle, sans laquelle l'égoïsme rationnel serait difficilement justifiable. Le réductionnisme, en revanche, fournit un argument logique au principe d'impartialité, qui jusque-là ne reposait que sur un impératif moral. Il est donc possible de l'utiliser dans un sens comme dans l'autre. Toute la question est de savoir dans quelle mesure. Parfit considère qu'un retournement ontologique aurait tendance à soutenir, davantage qu'à invalider, le principe utilitariste. Il reste néanmoins très prudent sur le sujet, refusant de cautionner l'utilitarisme sans condition. Ses arguments, développés dans le chapitre 15 de *Reasons and Persons*, reprennent (dans ses grandes lignes) ceux développés précédemment dans son article 'Later Selves and Moral Principles' (1973).

son plus grand bénéficiaire lorsqu'il sera adulte soit (2) pour le bénéficiaire d'une autre personne. Dans la mesure où l'épreuve est la même pour l'enfant, cela fait-il une différence qu'il en soit plus tard le bénéficiaire ou non ? Les non-utilitaristes affirment que cela fait bien une différence puisqu'il existe une vraie différence entre celui qui endure l'épreuve et celui qui bénéficie de ses bienfaits. Il est fondamental pour eux que celui qui souffre soit aussi celui qui bénéficie de la compensation de la souffrance, de la même manière qu'il est indispensable que celui qui a commis une faute soit également celui qui la répare. L'approche non-réductionniste permet plus aisément de défendre cette position<sup>13</sup>.

Le réductionnisme implique, au contraire, que l'enfant et l'adulte ne sont plus à proprement parler la *même* personne, mais qu'ils sont liés par la Relation R, qui peut être plus ou moins étroite selon les cas. Il est donc possible de supposer que l'enfant n'a pas de relation plus étroite avec celui qu'il sera adulte qu'il n'en a, par exemple, avec son jeune frère aujourd'hui. De ce point de vue, il n'est pas moins juste que son jeune frère profite de l'épreuve de l'enfant (par un don d'organe par exemple) qu'il en profite lui-même adulte. Pour le réductionniste, il n'y a pas de différence fondamentale entre les relations interpersonnelles et les relations intra-personnelles. Les principes de redistribution, qui s'appliquent généralement au sein de la même personne, peuvent désormais aussi s'appliquer entre les personnes.

Dès lors que l'on établit une analogie entre les relations entre personnes et les relations au sein d'une personne, on se doit d'appliquer également les principes de répartition aux unes et aux autres. Soit ils s'appliquent aux deux, soit ils ne s'appliquent à aucune. Les utilitaristes ne les appliquent nulle part. La raison en est, suggère Parfit, qu'ils adoptent une approche réductionniste (§114, 335). Il existe deux façons de ne pas accorder d'importance aux principes de répartition,

<sup>13</sup> Cela signifie-t-il pour autant que l'approche réductionniste est favorable aux utilitaristes ? Pas nécessairement. De la même façon que je peux dire qu'il est injuste qu'une autre personne que moi profite de mon malheur (ou bénéficie de mon travail), je peux également dire – étant réductionniste – qu'il est tout aussi injuste que son futur soi en profite. L'extension du champ d'application du principe utilitariste se fait au détriment de celui-ci. Considérons le point de vue réductionniste et supposons qu'il n'y a pas de différence fondamentale entre les relations entre personnes et les relations entre les segments temporels d'une même personne, autrement dit qu'il n'y a pas de raison de croire, par exemple, que l'enfant soit plus attaché à son moi futur qu'à son petit frère aujourd'hui. Si tel est le cas, l'enfant n'a pas davantage de raison de souffrir pour une personne qui lui est étrangère que pour son futur moi, et cela quelle que soit l'utilité potentiellement générée par son épreuve. Les gains d'utilité ne sont non seulement plus transférables entre les personnes mais ils ne le sont plus également entre les différentes périodes de vie d'une même personne, rendant toute maximisation d'utilité impossible. De ce point de vue, le réductionnisme ontologique est contraire à la philosophie utilitariste (§113, 334).

soit on ne leur accorde aucune portée, soit on ne leur accorde aucun poids. Rawls et Gauthier considèrent que, en étant non-réductionniste et en assimilant les choix d'une société aux choix d'un homme, l'utilitarisme ôte toute portée aux principes de répartition. Pour Parfit, au contraire, s'il est vrai que l'utilitarisme rejette tout principe de répartition, c'est parce qu'il refuse de leur donner du poids et non de l'étendue. Et cette explication-là est conforme à la théorie réductionniste de l'identité personnelle. Si, étant utilitariste, j'adopte le point de vue réductionniste, je serais forcé d'accorder *d'avantage de portée* mais *moins de poids* aux principes de répartition.

Expliquons-nous. Les utilitaristes pensent que les gains et les pertes d'utilité peuvent s'équilibrer ou se compenser, y compris entre différentes personnes. Autrement dit, l'augmentation du bien-être d'un ou de plusieurs hommes peut *moralement* compenser la souffrance d'un ou de plusieurs autres. Cette affirmation est naturellement très contestée. Rawls, en particulier, refuse d'égaliser les gains et les pertes d'utilité entre personnes afin de pouvoir en identifier le solde net. Parfit appelle cela *l'Objection comptable* ('Objection to balancing'). Cette objection repose, en partie, sur l'argument selon lequel il n'est pas possible de compenser les pertes d'utilité d'une personne par les gains d'une autre. C'est ce que Parfit appelle *l'Argument de la Compensation* ('the Claim about Compensation'). La différence entre les deux est tenue. L'objection comptable repose essentiellement sur une vision non-réductionniste de l'identité. Si les gains et les pertes d'utilité peuvent être comptabilisés ensemble au sein de la vie d'une même personne, ils ne peuvent pas l'être entre différentes personnes précisément parce qu'elles sont différentes. Cette objection n'est plus valable si l'on est réductionniste. L'argument de la compensation, lui, n'est pas contestable, y compris d'un point de vue réductionniste. La souffrance d'un individu ne peut pas être *moralement compensée* par le supplément de confort d'un autre, sauf peut-être dans le cas des personnes que l'on aime.

Que se passe-t-il alors si, étant utilitariste, j'adopte une approche réductionniste de l'identité personnelle ? L'argument sur la compensation, ainsi qu'on vient de le voir, tient toujours, mais sa portée et son poids sont, eux, modifiés. Le principe, tout d'abord, est étendu aux différents 'mois' temporels de la personne. Dans le cas déjà cité, il n'est donc plus justifié de faire souffrir l'enfant jeune pour le seul supplément de confort de l'adulte qu'il deviendra. L'enfant et l'adulte doivent être considérés comme deux personnes, sinon distinctes, au moins faiblement psychologiquement connectées. Les possibilités de compensation sont donc réduites, et la portée de l'argument l'est proportionnellement. A cela s'ajoute, le poids de ce principe qui est également diminué. La force de l'argument sur la compensation repose sur l'importance morale de l'identité personnelle. Mais, puisque le réductionnisme stipule que la vérité de la notion d'identité est

moins profonde ('deep'), alors on est en droit de conclure qu'elle est également moralement moins importante. Donc, le réductionnisme diminue à la fois la portée et l'importance de l'argument sur la compensation. Comme l'objection comptable repose en partie sur cet argument, celle-ci s'en trouve par conséquent également affaiblie (§115, 338).

Bien qu'il soit encore plus prudent sur le sujet en 1984 qu'il ne l'était déjà en 1973, la conclusion de Parfit reste essentiellement la même : l'utilitarisme est davantage plausible dans une perspective réductionniste que dans une perspective non-réductionniste. Son raisonnement peut, cependant, être critiqué sur au moins trois points. Tout d'abord, si la théorie utilitariste est plus cohérente avec le réductionnisme que sans, on ne peut pas en conclure qu'il constitue un meilleur argument pour l'utilitarisme que pour les alternatives à l'utilitarisme (Scheffler, 1982). Les philosophies morales n'étant généralement pas explicitement fondées sur une approche ontologique particulière, elles sont toutes dans une certaine mesure compatible avec le réductionnisme. Si l'utilitarisme semble mieux se défendre dans un cadre réductionniste que dans un cadre non-réductionniste, cela ne signifie pas *a priori* que les théories non-utilitaristes se défendent nécessairement mieux avec le non-réductionnisme. On pourrait ainsi très bien montrer que le perfectionnisme se justifie beaucoup mieux dans une perspective réductionniste (Schultz, 1986). Il est donc erroné de conclure que puisque 1° le réductionnisme est plus plausible que le non-réductionnisme et 2° que l'utilitarisme est rendu plus plausible par le réductionnisme, alors 3° l'utilitarisme est une théorie morale plus plausible que les autres.

À la suite de la publication de *Later Selves and Moral Principles*, Parfit a aussi été beaucoup critiqué sur l'absence de véritable justification de son principal argument, à savoir le fait que le réductionnisme rendant la notion de l'identité personnelle moins 'profonde', son importance morale dans la critique de l'utilitarisme s'en trouvait nécessairement réduite. Norman Daniels a été le premier à s'interroger sur la signification de la 'profondeur' de la notion d'identité chez Parfit, et le lien logique qui unirait fonctionnellement 'profondeur' et 'importance morale' : « This central assumption of Parfit's brilliant article is never argued for. My worry is not simply that there is vague, metaphorical language about the "depth" of certain facts and the "weight" of related principles. Rather, I am worried that there is not a single argument that reducing the (metaphysical?) depth of the fact is sufficient reason to warrant reducing the (moral) weight ascribed to a relevant principle. » (Daniels, 1979, 269). Parfit lui répond (indirectement) dans son article 'Personal Identity and Rationality' (1982), puis dans *Reasons and Persons*. Dans l'article de 1982, il considère deux types d'argument. Le premier, (A), peut se formuler ainsi : « Lorsqu'un fait rationnellement significatif est considéré comme

étant moins profond ('deep'), ou comme ayant moins d'implication, on peut vraisemblablement affirmer qu'il est moins significatif ; dans un cas extrême, il devient même possible d'affirmer qu'il a moins d'importance. » (Parfit, 1982, 230). Le second, (B), affirme que « Quand une relation importante tient à un moindre degré, il n'est pas irrationnel de penser qu'elle est moins importante » (Parfit, 1982, 232). Dans le cas d'une approche réductionniste de l'identité personnelle, (B) constitue naturellement un meilleur argument que (A), mais reste néanmoins insuffisant. Tout d'abord, il n'y a aucune raison *a priori* pour qu'il existe une proportionnalité entre l'affaiblissement d'une relation – qui est importante – et la diminution de son importance. Je peux ainsi être très faiblement connecté à mon futur moi, et néanmoins reconnaître l'importance de cette relation, aussi ténue soit-elle. Ensuite, dans les deux cas (A et B), rien n'est dit sur l'importance respective du fait ou de la relation en question par rapport aux autres faits et relations. A supposer que l'on découvre que l'identité est moins significative qu'on le pensait (A), ou que la Relation R tient à un degré plus faible qu'on ne le pensait (B), et que son importance est diminuée d'autant, cela ne signifie pas qu'elle devient *moins* importante que d'autres faits ou relations.

Troisièmement, on peut regretter que Parfit, dans sa présentation de l'utilitarisme, donne une vision partielle (voire partielle) de l'approche réductionniste (ou, diront certains, qu'en se référant au réductionnisme, il donne une vision partielle de l'utilitarisme). Le réductionnisme, tel que Parfit le présente, ne nie pas la possibilité d'une continuité de la personne dans le temps. Il stipule simplement que, contrairement à ce qu'affirment les non-réductionnistes, l'identité personnelle est une affaire de degrés. Je ne suis pas la même personne que celle que je serai dans le futur (éloigné) mais je ne lui suis pas non plus complètement étrangère (§102, 308). L'utilitarisme soit considère le groupe à l'image de l'individu (argument de Gauthier), soit ne considère pas l'individu comme une entité en soi (argument de Rawls). Parfit leur oppose un utilitarisme qui considère l'individu comme une multitude (Ferey, 2011). Sa justification de l'utilitarisme repose sur ce qu'il appelle lui-même la 'désintégration partielle' de la personne (§114, 336). Plus exactement, Parfit considère qu'en élargissant la portée du principe de répartition, les utilitaristes en ont significativement affaibli l'importance, rendant *in fine* leur approche *impersonnelle*. Cette version 'atomistique' de l'utilitarisme ne se justifie qu'au prix d'une distorsion majeure du réductionnisme, qui ne pose jamais le principe d'indépendance des mois temporels de la personne. On peut même aller plus loin, et affirmer avec Scheffler, que si le réductionnisme se

réduit effectivement à cela, alors l'utilitarisme – en tant que théorie morale – devient inopérant<sup>14</sup>.

S'il est possible d'accepter le réductionnisme et de rejeter la Proposition Extrême, alors on peut penser que le réductionnisme n'est pas synonyme de 'désintégration' de la personne. Parfit se situe-t-il dans cette configuration ? Oui et non. Parfit rejette effectivement la Proposition Extrême, mais on peut penser qu'il ne le fait pas pour les 'bonnes' raisons. Tout d'abord, il fait appel à l'argument de l'évolutionnisme pour justifier notre comportement prudentiel (§102, 308). Ensuite, il attribue la force de la Proposition Extrême à la prévalence de l'idée reçue selon laquelle l'identité personnelle serait un 'deep further fact', sans lequel nous n'aurions aucune raison de nous soucier de notre futur (§102, 311). Bien qu'il entrevoie très bien le fait que le réductionnisme n'implique pas nécessairement la Proposition Extrême, Parfit se montre réticent à l'idée de développer cette ligne théorique. Cela n'est pas, en soi, très surprenant. Rappelons, en effet, que la Proposition Extrême constitue à ses yeux 'la réfutation ultime' de l'égoïsme rationnel (§102, 312). Il adopte donc, semble-t-il, une position intermédiaire assez ambiguë qui se résume toute entière dans l'ambiguïté de l'expression 'désintégration *partielle*'.

## Conclusion

J'ai essayé dans le présent article d'évaluer l'impact de l'approche réductionniste de l'identité personnelle, introduite par Parfit, sur notre conception de la rationalité. J'ai montré, tout d'abord, que le réductionnisme pouvait remettre en question le canon de neutralité temporelle défendu par Sidgwick, mais également celui de l'actualisation à

---

<sup>14</sup> "For while I really have no idea what moral demands can most plausibly be made of atoms or events, it seems to me very doubtful that the demands embodied in standard forms of utilitarianism are likely candidates. Standard utilitarianism, by virtue of its requirement that agents devote energy and attention to their own projects and commitments in strict proportion to the weight in the impersonal calculus of their doing so, is rightly thought to be among the most demanding of moral conceptions. Because it is so demanding, standard utilitarianism has as much need as any familiar moral conception does for a plausible theory of character development and long-term motivation. That is, standard utilitarianism needs an account of the values, characteristic dispositions, and patterns of motivation which a person must have in order to live up to the rather rigorous standards it sets." (Scheffler, 1982, 237-38). La position de Scheffler doit, cependant, être nuancée. Tout d'abord, il joue d'une ambiguïté qu'il entretient savamment dans tout son article, et il est souvent difficile de faire précisément la part des choses entre les arguments qu'il oppose à Parfit et ceux qu'il oppose à une possible interprétation de Parfit. Cela complique inutilement les choses dans la mesure où la position de Parfit sur le réductionnisme manque également de clarté. D'une façon générale, l'argument de Scheffler est valable si Parfit adopte une version radicale du réductionnisme, autrement dit si celui-ci accepte la possibilité de la Proposition Extrême (Schultz, 1986).

taux constant généralement utilisé aujourd'hui en économie dans les modèles de choix inter-temporels. J'ai montré, ensuite, que le recours à la Relation R plutôt qu'à une conception dualiste de l'identité personnelle peut constituer un élément décisif dans l'invalidation du dogme de l'égoïsme rationnel et dans la construction d'une véritable théorie de la bienveillance rationnelle. L'approche réductionniste de l'identité personnelle peut donc avoir des conséquences extrêmement importantes dans la manière dont nous concevons la rationalité individuelle. Les économistes ont, par conséquent, de bonnes raisons de se tourner vers la philosophie, y compris vers la métaphysique. Et les travaux de Parfit constituent, de ce point de vue, une référence absolue. Le but de cet article est précisément de les inviter à s'intéresser davantage à son œuvre. L'espace nécessairement limité d'un article ne m'a malheureusement pas permis de rendre correctement compte de la complexité et de la subtilité de sa pensée, et j'ai été parfois obligé de simplifier ses vues. La question des biais temporels mériterait ainsi d'être développée et d'intégrer notamment ses vues sur ce qu'il appelle le biais envers le proche, et que nous appelons l'actualisation hyperbolique. Mais il me semble difficile de pouvoir le faire sans développer un minimum les biais vers le futur et le présent, notre rapport au passé, et la rationalité ou l'irrationalité de certains désirs temporels. De plus, ces considérations sont – chez Parfit – traitées indépendamment de la question de l'identité personnelle qui nous intéresse ici. Je renvoie donc, pour l'heure, le lecteur à la littérature existante (Dougherty, 2011, Greene and Sullivan, 2015, Hedden, 2015).

## Références bibliographiques

- Brink, David O. 2011 Prospects for Temporal Neutrality. In Craig Callender (ed.), *The Oxford Handbook of Philosophy of Time*. Oxford and New York: Oxford University Press, 353-381.
- Daniels, Norman. 1979. Moral Theory and the Plasticity of Persons. *The Monist*, 62(3): 265-287.
- Davis, John B. 2003. *The Theory of the Individual in Economics : Identity and Value*. London: Taylor & Francis.
- Davis, John B. 2010. *Individuals and Identity in Economics*. Cambridge: Cambridge University Press.
- Dougherty, Tom. 2011. On Whether to Prefer Pain to Pass. *Ethics*, 121(3): 521-537.
- Ferey, Samuel. 2011. Paternalisme libéral et pluralité du moi. *Revue Économique*, 62(4): 737-750.
- Greene, Preston and Meghan Sullivan. 2015. Against Time Bias. *Ethics*, 125(4): 947-970.
- Hedden, Brian. 2015. *Reasons without Persons : Rationality, Identity, and Time*. Oxford: Oxford University Press.

- Johnston, Mark. 1997. Human Concerns without Superlative Selves. In Jonathan Dancy (ed.), *Reading Parfit*. Oxford: Blackwell Publishers, 149-179.
- Nagel, Thomas. 2012. *Mortal Questions*. Cambridge: Cambridge University Press.
- Olson, Eric. 1997. *The Human Animal: Personal Identity Without Psychology : Personal Identity Without Psychology*. New York: Oxford University Press.
- Parfit, Derek. 1982. Personal Identity and Rationality. *Synthese*, 53(2): 227-241.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Parfit, Derek. 1999. Experiences, subjects, and conceptual schemes. *Philosophical Topics*, 26(1-2): 217-270.
- Scheffler, Samuel. 1982. Ethics, Personal Identity, and Ideals of the Person. *Canadian Journal of Philosophy*, 12(2): 229-246.
- Schultz, Bart. 1986. Persons, Selves, and Utilitarianism. *Ethics*, 96(4): 721-745.
- Shoemaker, David. 1997. Parfit on Identity. In Jonathan Dancy (ed.), *Reading Parfit*. Oxford: Blackwell, 135-148.
- Shoemaker, Sydney and Richard Swinburne. 1984. *Personal Identity. The Dualist Theory*. Oxford: Basic Blackwell.
- Wachsberg, M.M. 1983. *Personal Identity, the Nature of Persons, and Ethical Theory*: University Microfilms.
- Williams, Bernard. 1970. The Self and the Future. *Philosophical Review*, LXXIX(2): 161-180.
- Williams, Bernard. 1976. *Problems of the Self: Philosophical Papers 1956-1972*. Cambridge: Cambridge University Press.







# Economics of paternalism: the hidden costs of self-commanding strategies

Christophe Salvat

► **To cite this version:**

Christophe Salvat. Economics of paternalism: the hidden costs of self-commanding strategies. Journal of Philosophical Economics, 2015, IX (1), pp.102-124. <<http://www.jpe.ro/?id=b1>>. <halshs-01233100>

**HAL Id: halshs-01233100**

**<https://halshs.archives-ouvertes.fr/halshs-01233100>**

Submitted on 24 Nov 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

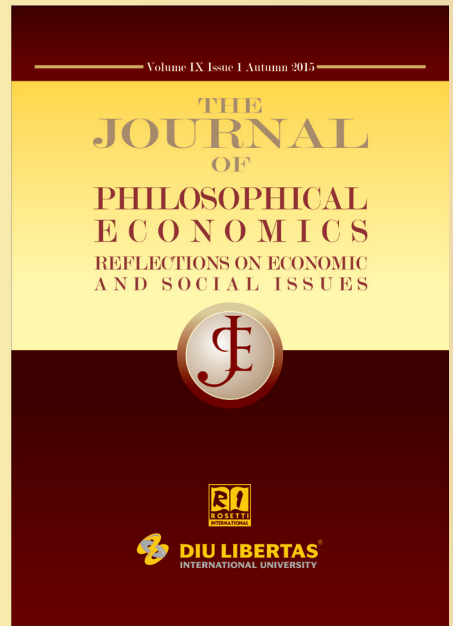
THE JOURNAL OF  
PHILOSOPHICAL ECONOMICS:  
REFLECTIONS ON ECONOMIC  
AND SOCIAL ISSUES

Volume IX Issue 1 Autumn 2015

ISSN 1843-2298

Copyright note:

No part of these works may be reproduced  
in any form without permission from the  
publisher, except for the quotation of brief  
passages in criticism.



*Economics of paternalism: the hidden  
costs of self-commanding strategies*

**Christophe Salvat**



# Economics of paternalism: the hidden costs of self-commanding strategies

Christophe Salvat

**Abstract:** The paper proposes an economic assessment of paternalism by comparing different alternative responses to dynamically inconsistent behaviors consecutive to hyperbolic discounting. Two main types of action are possible, self-commanding strategies and paternalism. The first category includes personal rules and pre-commitment. The second can be subcategorized between coercive and non-coercive forms of paternalism, which are respectively associated (although it is debatable) with legal paternalism and with ‘nudges’. Despite being self-inflicted, self-commanding strategies are actually not cost free and can result in a dramatic cutback of people’s freedom of choice. Likewise, legal paternalism can, on occasion, be less harmful than personal rules or pre-commitment; similarly, nudges can be more invasive and less effective than their proponents want us to believe. The aim of this paper is not to propose any standardized form of response to irrational behavior (whatever that may mean) but to argue, on the contrary, that every case should be individually appraised. Individual situations can be remedied by self-commanding strategies or by paternalistic policies, either in isolation or in combination.

**Keywords:** libertarian paternalism, personal rules, nudges, self-confidence

## Introduction

The issue of paternalism has become popular amongst social thinkers over the last decade, and more specifically amongst economists. The success of Sunstein and Thaler's 'Nudge' theory is certainly the best illustration of this revival. Nudge first capitalized on three decades of behavioral studies demonstrating a lack of rationality in individuals and the inability of economic theory to account for this. Economists have found that libertarian paternalism provides them with a way to increase individual rationality whilst preserving consumer sovereignty and thus, more importantly, they found a way to reassert the validity of their theoretical standard of rationality. The philosophical foundations of this theory have often been challenged; however, its economic basis has not yet been thoroughly examined. Whilst Sunstein and Thaler have little difficulty in demonstrating that nudges can have positive effects, they have not yet managed to successfully show that nudges are the most efficient form of paternalism, or that paternalism is even the best response to people's imprudence. The present article addresses this issue and claims that nudges are only one alternative amongst many, and that a careful cost-benefit analysis is needed before generalizing this kind of action.

This paper is structured around six sections. The first section narrows the discussion to very specific types of imprudent actions, those generated by variable time discounting rates. This allows us to avoid questioning the rationality of individuals' preferences, and to restrain the possible field of paternalistic actions to actions that individuals themselves would prefer not to undertake. It does not necessarily mean that paternalist or non-paternalistic interventions are not justified in other cases or that they are always justified in that type of case. Limiting the scope of the discussion to a commonly accepted type of irrational behavior is solely a way of focusing on the comparative advantages of alternative responses to irrationality (whatever irrationality actually means). The second section presents the first possible response to imprudence, i.e. self-commanding strategies, which either refer to personal rules or to pre-commitment. The third section discusses the possibility of paternalistic actions, which can either refer to allegedly coercive actions (legal paternalism) or allegedly non-coercive actions (nudges). The fourth section compares the relative cost (in terms of freedom of choice) of these possible solutions. The last section concludes this study by

rejecting pro and anti-paternalism supporters back-to-back and suggests a differentiated approach based on pragmatism.

## Opportunities of paternalism

In this section I consider the conditions necessary to make paternalism justifiable. There are many approaches to this question. One consists of stating that paternalism is never justifiable, however large the benefits and however low the cost might be for the person. This is a position that I shall not consider here since my aim is to assess whether, and which, cases of paternalism can be economically justifiable. I do not intend this study to be a defense of paternalism. I would rather prefer to condemn paternalistic actions on their relative inefficiency and/or cost than for pure ideological reasons. A second approach consists of saying that paternalism is justifiable whenever it is beneficial to the people interfered with. Although partisans of paternalism rarely explicitly endorse this view, I believe it is the most common amongst them. One can argue the belief that benefits should always be related to their potential cost, whether immediate or long term. And the main cost of paternalism is, as we know, the reduction of personal autonomy. There might be other negative consequences, but I shall only consider this one in the present study.

There are two ways of assessing the benefits of paternalism. The first is to objectively claim that individuals would be better off with an external intervention than without and the second is to trust individuals in their own appreciation of what is good for them. Economists have a strong preference for the second one, and it is often due to this point of view that paternalism is discarded as being totalitarian. Both approaches have their own merits and weaknesses. Individual preferences are not always 'rationally acceptable' even if they respect the conditions of completeness and transitivity set by economists, and we do not all share a same 'objective' view of individual well-being. Considering both positions, I believe the preference approach to be more reasonable than the objective approach. I shall therefore adopt here the standard economic view, according to which individuals are rational in so far as they satisfy their own preferences (even if I am aware that *some* preferences are intrinsically irrational). A second reason motivates this choice: partisans of an objective approach to well-being are usually more sympathetic to the idea of

paternalism than the defenders of the subjective approach. If I were to show that paternalistic actions can be more efficient and/or less costly than non-paternalistic actions, my point would be even stronger had I demonstrated it from a traditionally anti-paternalistic point of view.

Let us now consider cases which could *potentially* qualify for paternalistic actions. My main concern here is to identify potential cases of inconsistent behavior. I propose to focus here only on one kind of inconsistency, preference inconsistencies over time or Dynamically Inconsistent Behavior (DIBs).

Dynamically Inconsistent Behaviors are extremely common in the population and are often caused by hyperbolic discounting. DIB's have been associated with a wide range of expected utility anomalies from procrastination (Akerlof, 1991; O'Donoghue and Rabin, 2001) to credit-card debt (Laibson, 1997) or addiction (Gul and Pesendorfer, 2007; Herrnstein and Prelec, 1992). Discounted-Utility models are classically exponential, using a reducing factor of  $1 / (1 + k)^t$  where  $k$  is the constant discount rate and  $t$  is the length of the delay. Given a constant discount rate, the discounting becomes exponential. The value of future rewards directly (and exclusively) depends on the length of time that individuals have to wait for them. Unless the discounted value of rewards expected in  $t_n$  exceed the non-discounted value of present rewards, individuals are considered perfectly rational in preferring present to future benefits. Preferring the present to the future, therefore, is a matter of personal taste rather than an inconsistent inter-temporal choice. Exponential discounting models *a priori* exclude any possibility of paternalism as they assume the preference for the present is voluntarily chosen. Any attempts to prevent a person's future harm is not only illegitimate (as it violates individual freedom) but also irrational (as it does not maximize utility). This is the position adopted by Gary Becker in his works on addiction for instance (Becker and Murphy, 1988). On the contrary, discounting rates are varying in cases of hyperbolic discounting. The expression 'hyperbolic discounting' is actually slightly misleading since time discounting functions are not strictly hyperbolic but quasi-hyperbolic (Laibson, 1997). Their first axiomatic analysis was presented by Robert Strotz in 1955-56 (Strotz, 1955).

$$U^t(u_t, u_{t+1}, \dots, u_T) = \delta^t u_t + \beta \sum_{\tau=t+1}^T \delta^{\tau-t} \delta^{\tau} u_{\tau}$$

Where  $0 < \delta, \beta \leq 1$

If  $\beta = 1$ , then exponential discounting preferences

If  $\beta < 1$ , then (quasi) hyperbolic discounting preferences

In hyperbolic discounting the bias is introduced by the sense of proximity of the delay (and not just its duration). The closer it gets the higher the time preferences go. Hyperbolic discounting functions use a multiplying factor  $\beta < 1$ , reducing discounting rates as the discounted event moves further away in time. Events located in a near future are discounted at a higher discount rate than events located in a distant future. Future rewards are reduced by a factor of  $1 / (1 + kt)^{\beta\alpha}$  where  $\alpha$  and  $\beta$  are greater than zero. The longer is the delay, the lower is the discount rate. Alternatively, the closer is the expected reward, the more discounted is the expected event. What makes this bias so specific (and dynamically inconsistent) is that a person's rate of discounting increases as the deadline approaches. As a matter of fact, discount rates change with the passage of time, as individuals reevaluate their plans accordingly, abandoning the plans they made in the past and making new plans for the future. Plans that had been rationally elaborated in the past with a long term discount rate cease to be optimal once the deadline approaches and a (higher) short-term discount rate is applied. Periodically reevaluating one's plans logically leads to long-term utility washouts. The latest plans are, in that respect, irrational.

More interestingly for our purpose, the closer the deadline gets, the less desirable the next option becomes, to the point where the last one is reluctantly chosen by the agent. Take the following example presented by O'Donoghue and Rabin. 'Suppose you usually go to the movies on Saturdays, and the schedule at the local cinema consists of a mediocre movie this week, a good movie next week, a great movie in two weeks, and (best of all) a Johnny Depp movie in three weeks. Now suppose you must complete a report for work within four weeks, and to do so you must skip the movie on one of the next four Saturdays. When do you complete the report?' (O'Donoghue and Rabin, 1999: 109). The best plan, all things considered, is naturally to complete your report the first week and to only miss the mediocre film. But, given your time-biased preferences, you



prefer going to the cinema the first week and commit to write the report the following week. Being time-biased (and being completely naïve about it, I shall come back to that later), you choose to indulge on the first Saturday off (even if it is to see a mediocre movie) and commit yourself to write the report in the following week (in order to be still able to see the great movie and the Johnny Depp movie, which has your preference). The following week, however, your discount rate having increased, you logically reconsider your previous choice, and go to the cinema again. The same thing happens in week 3 obliging you to complete the report on the last Saturday and to miss Johnny Depp's latest movie. Missing Johnny Depp's latest movie to write a report is clearly not a voluntary choice.

To fully appreciate the effects of time inconsistency, one must therefore take in consideration a further fact: sophistication. Whereas some people are fully aware of their time inconsistency (due to their past experiences for instance), others seem genuinely taken aback each time it happens. Strotz initially distinguished two types of individuals, the 'spendthrift' individuals who do not recognize their dynamic inconsistency and the 'thrift' individuals who do (Strotz, 1955). They are now more commonly referred as 'naïf' and 'sophisticate'. Let us keep the latter terminology. The distinction is an important one. In terms of welfare, naïfs usually suffer from bigger utility losses than sophisticates. While sophisticates (even incorrectly) anticipate their preference reversal, and (eventually) take action using self-control or external commitment device (see next section), naïfs repeatedly choose dominated options, endlessly incrementing their utility losses. Even a small hyperbolic bias ( $\beta$  close to 1) can have dramatic effects on the naïfs' long-term welfare, whilst they are theoretically harmless for sophisticates. Paternalism, consequently, should primarily aim at stimulating sophistication. I strongly emphasize the fact that, contrary to common beliefs on imprudence or irrationality, the absolute magnitude of the person's discount rate has no bearing *per se* on her utility loss. She is not imprudent because she discounts too much her future rewards but because she discounts them more as they get closer. In fact, over long periods, hyperbolic discounting functions generally tend to discount future rewards at a lower rate than exponential ones.

Differentiating naïfs from sophisticates can be useful to assess the legitimacy of paternalistic actions [11]. It is indeed easier to identify a person's real preferences when she is sophisticated rather than naïve. Even if they fail to satisfy their initial

preferences, the fact that sophisticates expected a future preference reversal and took step against it is a testimony of their real preferences. This means that their action is involuntary, and henceforth eligible to paternalistic policy. Such is not the case of naïfs, who systematically deny their initial and utility-maximizing choice when they fail to implement it. When it comes to paternalism, the result is puzzling. On one hand, unlike sophisticates, naïfs are badly hurt by small hyperbolic bias. On the other hand, paternalistic actions can be better justified in cases where individuals are sophisticated than naïve. That ironically makes paternalism most legitimate when unnecessary.

## Self-commanding strategies

In the following, I shall discuss the potential responses to Dynamically Inconsistent Behaviors (DIBs): paternalistic and non-paternalistic policies. This will then enable me to respectively compare their relative cost. In the present section, I start with self-imposed policies only accessible to sophisticated agents.

One of the most common liberal beliefs is that it is always better to let individuals choose for themselves. Let us suppose for now that individuals are sophisticated. Despite being spontaneously inclined to overvalue earlier rewards relative to later ones, sophisticated agents can show a remarkable economic rationality. A first way to deal with DIBs is to adapt. This is the strategy of consistent planning initially described by Strotz. Consistent planning is a bargaining game, played between the present and future selves of a single agent. Plans that her present self knows will be overturned by her later self are definitively discarded so that a subgame equilibrium (known as the Strotz-Pollak equilibrium) is to be found amongst the remaining feasible options (Pollak, 1968; Peleg and Yaari, 1973; Goldman, 1980). Despite being freely chosen, this solution remains highly unsatisfactory since it does not really address the underlying issue.

Two other options can be considered, personal rules and pre-commitments. They constitute the main self-commanding strategies used to counteract DIBs. While personal rules are purely self-commanded, pre-commitments need the mediation of external individuals. Let us first consider intermediated self-commanding strategies.

Intermediated self-commanding strategies usually take the form of a contract. Ulysses demanding his crew men to bind him to the ship's mast when approaching the sirens is a classic example (Elster, 1979). There are many other ways to trump temptations, many of which do not involve physical constraints. Schelling listed a number of them, among which relinquishing authority to somebody else, committing or contracting, disabling or removing yourself, removing mischievous resources, submitting to surveillance, incarcerating yourself, arranging rewards or penalties, rescheduling your life and setting yourself the kinds of rules that are enforceable (Schelling, 1984: 6-7; Schelling, 1992). But the most common cases either rely on a strong moral binding, like a promise, or on a short-term game plan like a pre-agreed utility cost (Gul and Pesendorfer, 2001; Gul and Pesendorfer, 2004). Deadlines (and the moral or financial sanctions that go with them) are notoriously useful for coping with procrastination. Similarly, gyms do not thrive in spite of, but because of, their rigid and expensive membership terms. Note, however, that third parties who are implicated within these self-imposed rules as part of the enforcement mechanism are not necessarily mandated for this task and are sometimes not even aware of their mediation role. Social blame and public reprobation can be a deterrent as effective as any financial penalty. This is the strategy adopted by members of Alcoholics' Anonymous for instance.

Pre-commitments are usually highly effective but they present at least two important drawbacks: they require foresight and they lack flexibility. Like any self-commanding strategies, pre-commitments are only open to sophisticated individuals. This, in itself, considerably reduces their scope. But, in addition to that, they are necessarily restricted to expected events. Ulysses escaped his fate because he precisely knew where the sirens would be. Had he not known, he would have had to stay bound to his mast throughout the whole duration of the Odyssey. Knowing one is time biased is one thing, but knowing when one will need control is another. Let us assume that one knows when control will be needed, and that a contract has been accordingly agreed, there is still the possibility for the contract to be overly or insufficiently strict. If the contract is insufficiently strict, it will be ineffective. If the contract is excessively strict, then it can be unnecessarily harmful if not dangerous. Consider, for instance, the case of removing 'mischievous' resources, discussed by Schelling (Schelling, 1984). A woman expecting her first child has decided to fully and consciously experience the birth of her baby: she wants to deliver without any anesthetic. The physician proposes to have an anesthetic ready for her to use in case she needs it. She knows, and the experience of

others show (Christensen-Szalanski, 1984), that, with the prospect of the pain drawing nearer, she will reconsider her choice, and ask for pain relief, a choice that she thinks she will regret later when the baby is born. She therefore refuses the physician's proposition and demands all anesthetic to be removed from the room when she gives birth. If the physician keeps the anesthetic close at hand, it is more likely that she might use it. But if the physician does not offer her the possibility of using it at all, then she might experience an unbearable pain, and never want to give birth to another child ever again. Pre-commitments are instrumental in our well-being but it is irrational to use them at any cost.

Non-mediated self-commanding strategies, commonly known as personal rules, offer more flexibility. Personal rules can be implemented without any external help and give (at least at first sight) sophisticated agents more latitude to deal with their own inconsistencies. Individuals typically implement personal rules to stop smoking, avoid procrastination, go on a diet or work harder, all problems raised by DIBs. Unlike pre-commitments, personal rules are often kept quiet, either because individuals are ashamed of their problems or because they are ashamed that they are not able to solve them and could appear weak. Note that when personal rules are made public, it is generally to use public reprobation or personal shame as an additional incentive to obey one's rule. Personal rules can thus easily turn into pre-commitments. When they are kept secret, and nobody but the individual can see that they are being broken, enforcing them becomes a real challenge. Diets are suspended on bank holidays, then on birthdays, then on week-ends, and then every time one goes to the restaurant or one has a pizza. Similarly, cigarettes are first tolerated in stressful times, and end up being smoked on all occasions. For having witnessed it around us, and for having personally experienced it, we all know that personal rules, when they are kept personal, do not generally work.

To be effective, personal rules require willpower. Willpower is, by definition, the ability to resist temptation (Ainslie, 1992: 142-143). This is, as we all know, rather unequally distributed among the population. Some people are lucky enough to show a startling level of willpower (or so it seems), when others are victims of their incapacity to resist natural impulses. And there are many ways to improve one's willpower (Holton, 2009; Bénabou and Tirole, 2002; Bénabou and Tirole, 2004). But appearances can be also deceptive: seemingly strong-willed individuals can also turn out to be weak

but severely self-constrained. Willpower is commonly believed to be a natural feature. It may well be to some extent. But it is mostly acquired through experience, education and good faith. The economists Roland Bénabou and Jean Tirole showed that willpower essentially works as a *reputational capital*, growing each time a personal rule is successfully tested and decreasing each time it fails the test (Bénabou and Tirole, 2004). Individuals are not obliged to test their willpower, in which case they indulge their current impulses without losing confidence in their willpower. For Bénabou and Tirole, individuals who already have a certain reputational capital to preserve have therefore an additional incentive to comply with their personal rules, making strong-willed individuals even more likely to stick to the rules. Conversely, weak-willed individuals who have difficulty in resisting temptation are getting weaker and weaker as their rules lapse. Two factors explain this vicious circle. Firstly, the smaller an individual's reputational capital, the smaller their incentive to protect it. Secondly, once a rule has been broken, it significantly loses its moral power. Lapsing once to a rule creates what Ainslie calls a 'precedent', and makes any further lapse easier to happen [2]. Partners, for instance, usually take a vow not to cheat on each other. The vow is explicit for married couples, but is usually implicit for unmarried couples. When opportunities to cheat turn up (as they often do), they are usually dismissed thanks to the moral cost (the shame) that would inevitably result from breaking this personal engagement. But once one partner has cheated then the harm is done and there is no further disutility in cheating a second time, and then a third etc. The (moral) cost of breaking the rule is mostly entirely born with the first lapse. Cheating again only marginally adds to the guilt felt the first time. According to this logic, strong-willed individuals naturally tend to get stronger when weak-willed ones naturally tend to get weaker. This is, however, not necessarily the case. Naturally weak-willed individuals can also turn out to be the most rigid rule followers.

This claim can seem surprising at first. Psychologist George Ainslie brilliantly demonstrated that precisely because individuals were weak-willed or believed themselves to be so, they were prone to take drastic steps to avoid future failures. If they want to resolve their problems alone, weak-willed individuals must be uncompromising with themselves. They usually start by denying themselves any excuse to lapse to the rule. People who systematically follow their rules and seem to be strong-willed may in fact be under confident and therefore uncompromising in their behavior. They are simply overly harsh with themselves, punishing themselves when they lapse,

or even worse, externalizing the punishment. This is what Schelling calls side-betting (Schelling, 1960). Side-betting differs from commitment since commitment requires the explicit consent from a third part, when side-betting does not. One way of externalizing punishment is typically to submit oneself to public reprobation or shame. This is the solution adopted by members of Alcoholic Anonymous. Some determined (yet weak-willed) individuals are more imaginative. Take for example the case of the drug addicted physician quoted by Schelling: 'In a cocaine addiction center in Denver, patients are offered an opportunity to submit to extortion. They may write a self-incriminating letter, preferably a letter confessing their drug addiction, deposit the letter with the clinic, and submit to a randomized schedule of laboratory tests. If the laboratory finds evidence of cocaine use, the clinic sends the letter to the addressee. An example is a physician, who addresses a letter to the State Board of Medical Examiners confessing that he has administered cocaine to himself in violation of the laws of Colorado and requests that his license to practice be revoked. Faced with the prospect of losing his career, livelihood, and social standing, the physician has a powerful incentive to stay clean.' (Schelling, 1992: 167). This case is exemplary of the extreme means that some people are willing to use to overcome their weakness. Ainslie was the first to demonstrate that a poor perception of one's willpower, or too high an image of other people's willpower, often leads weak individuals to use excessively harsh means. Men and women dread flexible rules as they dread their own weakness. This explains why some people prefer to stop socializing altogether when they go on a diet, or why others turn into workaholics to avoid procrastination. Compulsion succeeds to personal rules, and prudence gives way to miserly behavior and bigotry.

However, self-commanding strategies are not always that costly. They are still widely used with success or near success in everyday life. Everyone lapses now and then, to one's personal rules, but overall we rather successfully comply with them. And in the vast majority of cases, these lapses are not significant enough to call for extra sanctions. Did I eat a second *pain au chocolat* this morning when I only intended to eat one? Well, I'll try to be more cautious next time. That's it. Did I oversleep this morning? Well, I'll put an extra alarm clock on my phone for tomorrow. There is no need to pre-commit, or to venture into complex and harsh side-betting. Those strategies should be reserved for individuals who are absolutely unable to follow a rule, and to cases in which lapses have serious consequences that cannot be forgiven.

## Possibility of paternalism

In the previous section I presented non-paternalistic solutions to DIBs. Although many, if not most, cases of spontaneous responses to hyperbolic discounting can be resolved without external interventions (Holton, 2009), others, however, may require such action to be taken. The use of external intervention is normally related to cases where individuals are naïfs and therefore unaware of their dynamic inconsistency, or when the self-imposed solutions of sophisticated individuals are too costly in terms of autonomy. In these situations, paternalism might then be considered. Paternalistic actions can be characterized as unsolicited actions undertaken by a benevolent agent and motivated by someone else's irrational or imprudent behavior. They specifically target the decision-making process by checking or by boosting the other individual's motivation. This is either achieved by emotional or psychological tricks or, alternatively by means of authority. By unsolicited actions, I mean all actions that have not been explicitly called for. That naturally includes, but is not limited to, coercive actions. Unsolicited actions also include unsolicited advices or unsolicited help. Unsolicited actions are by definition not pre-consented, but they can be endorsed *a posteriori*.

Paternalism has been traditionally discarded as a harmful hindrance to individual freedom (Dworkin, 1983: 142-143; Feinberg, 1986; Marneffe, 2006). These authors were essentially concerned with the loss of individual freedom potentially resulting from paternalistic actions. This is indeed a major moral consideration, although actually not the only one (Salvat, 2014). Recently, Sunstein and Thaler argued that paternalism can be as innocuous as a nudge, hence opposing non-coercive paternalistic 'nudges' with traditional legal – and coercive – paternalism. Their argument is, however, somewhat misleading. It could be argued first that making a coercive action less visible or painful does not mean that the action is any less coercive. Manipulation can also be a form of coercion (Bovens, 2009). Paternalistic laws, secondly, can claim to be freely chosen in democratic countries. Paternalism can therefore be a collective choice, as Rawls himself admitted (Rawls, 1971). I therefore propose to judge paternalistic actions not as what they seem, but on what they actually cost to individuals in terms of freedom of choice. In the previous section, I showed that self-inflicted rules or pre-commitment could be costly; I want to see here how paternalistic actions compare to this.



How are we to assess the relative cost of paternalistic policies? As I stated at the beginning of this study, I am only interested here with paternalism's impact on individual freedom. So by cost one should understand infringement of freedom of choice. It is also clear enough by now that paternalistic policies are assessed relatively to non-paternalistic policies, which also have a cost (personal rules, pre-commitment, side-betting etc). Last, but not least, I consider that the severity (and therefore the cost) of paternalistic and non-paternalistic policies not only depends on the agent's actual willpower but also on the extent of her subjective (and often mistaken) appreciation of her own willpower. It is therefore important to distinguish willpower and self-confidence. I shall refer to willpower as the probability to comply with the rule previously adopted or the decision previously taken and use self-confidence to refer to the perception an agent has of her own willpower. Hence an agent is said to be over-confident when, for instance, she believes her probability of sticking to her rule or her decision is  $\frac{3}{4}$  when it is actually  $\frac{1}{4}$ . Conversely an agent is said to be under-confident when she believes the probability that she will stick to her rule is  $\frac{1}{4}$  when it actually is  $\frac{3}{4}$ . An agent who has an exact appreciation of her willpower is fully sophisticated.

I can now offer a (rough) typology of paternalistic actions that are made possible in the case of DIBs:

- Lack of foresight or bad luck ( $P_1$ )
- Lack of sophistication ( $P_2$ - $P_3$ )
  - agents fail to comply with their personal rules ( $P_2$ )
  - agents commit themselves to unreasonably harsh or demanding rules ( $P_3$ )
- Absence of sophistication ( $P_4$ )

Let me develop them in order.

$P_1$  refers to cases in which individuals, although sophisticated, are unable to resort to self-commanding strategies. Sophisticated agents, for instance, know that they are time inconsistent, but they fail to correctly identify situations in which their time bias will be an issue. This is a problem in particular for pre-commitments, which require complete and accurate anticipation. Consider the case of Ulysses. Ulysses knows where the sirens will be and is bound accordingly. He is free for the rest of the journey.

Consider, next, the case of Moby Dick's Captain Ahab, initially quoted by Schelling



(Schelling, 1984). In the film (this particular episode is not in the book), Ahab seriously injured his leg in the water. Back on the boat, Ahab is held down by his crewmen so that the blacksmith can cauterize the stump with a hot iron. Ahab refuses to be burned. Ahab is in no way different from Ulysses except in that that he did not expect his predicament. Had Ahab anticipated his accident, he would have instructed his crewmen in one way or the other beforehand. Personal rules are a more appropriate response to this kind of situation but, as illustrated by Ulysses' tale, they are not applicable to all cases.

$P_2$  and  $P_3$  refer to cases in which personal rules have been implemented but are either insufficiently effective or too costly. Let us first consider the case of insufficiently effective personal rules ( $P_2$ ). Failures to comply with personal rules are usually believed to be the result of weak willpower. There is actually more to that. What is really decisive is how well you appreciate the level of your willpower rather than the level of your willpower itself. Let me give you an example. Suppose there is a customer with a willpower level of  $1/4$ . She goes each day to the cafeteria, and each day she has the choice between fruit salad and cheesecake. Her preference, before going to the cafeteria, is always to take fruit salad. But having a level of willpower of  $1/4$  the probability of her taking the fruit salad is only one of every four occasions. Being sophisticated to some extent, she does not know her real chances but she knows that she has some chance of taking the cheesecake in the cafeteria. Imagine first that she overestimates her willpower. Suppose that she estimates her chances of choosing the fruit salad at  $3/4$ . Because she believes her chances to be already quite high, she gives herself a relatively soft but rather ineffective rule. Paternalistic solutions are then possible ( $P_2$ ). They are not necessarily justified (see next section) but there is also an opportunity here. Alternatively, the customer underestimates her willpower, and sets a personal rule that, although very effective, can be unduly costly (as illustrated by Schelling's cocaine addict physician). In those cases, paternalism can potentially be offered as a less coercive solution ( $P_3$ ).

**Table 1:** Possibility of paternalism

Self-confidence	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{8}$
Willpower	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
Personal rule	Insufficient	Adequate	Excessive
Paternalism (possibility of)	$P_2$	None	$P_3$

Finally,  $P_4$  refers to cases where self-commanding strategies are simply not an option due to individuals' naivety. This is a theoretically straightforward case that supporters of paternalism are eager to use. In practice, however,  $P_4$  raises a number of ideological issues. Apart from some particular cases, it is indeed debatable whether people can be entirely unaware of being time biased. Mentally deficient adults and young children undoubtedly belong to this category and need external guidance. For the rest of the population, I am personally inclined to believe that, although a lot of people are excessively over-confident, very few are completely naïve. The opposition introduced by Strotz between sophistication and naivety is fundamental but one would be misled to oppose pure sophisticated and pure naïve individuals. The vast majority of people actually belong to an intermediate category.  $P_4$  is nonetheless possible.

## Price of paternalism

Let me now assess the cost of paternalism to each of these cases separately.  $P_1$  and  $P_4$  are particular in the sense that they do not allow comparisons between paternalistic and non-paternalistic responses to hyperbolic discounting. Since it is my aim in this article to assess the cost of paternalism relative to self-commanding strategies, I shall exclusively focus on  $P_2$  and  $P_3$ .

Table 1 showed that paternalism is either made possible by excessive or insufficient self-confidence. It is now our responsibility to find out when these paternalistic actions can be actually justified. As I explained above, I believe paternalism is neither intrinsically legitimate nor illegitimate, but that its legitimacy depends on its relative cost. And by cost, I mean the reduction of freedom of choice involved by its implementation. It has been sometimes argued that paternalism does not necessarily

infringe people's autonomy, autonomy being understood as the expression of the person's real will. It is a defensible point of view but one that has no chance being heard by libertarian thinkers. I therefore propose to assess the relative impact of paternalistic and non-paternalistic solutions to hyperbolic discounting in terms of choice rather than autonomy.

I also argued in the previous section that the efficiency of personal rules depended not only on willpower, as commonly presumed, but more importantly on self-confidence. The higher the gap is between willpower and self-confidence, the harder the intervention needs to be. Individuals generally set themselves a rule whose severity is inversely proportional to their self-confidence. The more confident they are, the more innocuous is the rule they set for themselves. But, at the same time, the weakest is their willpower, the harder the rule needs to be. The difference represents the relative cost of paternalism (Table 2).

Table 2: Cost of Paternalism

Agents	A	B	C	D	E
Self-Confidence	1	$\frac{4}{5}$	$\frac{3}{5}$	$\frac{2}{5}$	$\frac{1}{5}$
Willpower	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{2}{5}$
Personal rule (severity)	Nil	Soft	Medium	Adequate	Too Hard
Intervention (relative cost of)	High	Medium	Low	Nil	Negative

Agent A thinks that she does not need a personal rule because she believes that she is strong enough to avoid any preference reversal. Yet being actually weak, she does need a rule. Because she failed to see that, the rule that might be imposed on her will be resented as extremely costly. B and C are also over-confident but they are sophisticated enough to adopt a personal rule. Because B's rule is too soft relative to her willpower, she might need – if she really wants to achieve her aim – additional self-commanding or paternalistic devices. Note that B and C do not necessarily need paternalism. In many cases, occasionally lapsing rules are of no great consequence. Furthermore, individuals progressively adapt their rules to their actual willpower. But if paternalistic actions are nevertheless undertaken, their relative cost should be

moderate. The more the severity of the paternalistic rule approaches the severity of the personal rule, the less it is resented as infringing one's freedom of choice. When they are equally severe, the relative cost of the paternalistic intervention is nil (D). This is what happens when the agent correctly evaluates her willpower. The paternalistic rule (the law for instance) then perfectly corresponds to the rule she would personally adopt. Despite being externally imposed on her, the law does not generate extra cost for her to follow. In other words, a same rule (say a law forbidding selling extra-large cups of sodas) can be resented as coercive by some people (like A) whilst others (like D) will not suffer from it. One can even go further as some people (like E) might actually see their freedom of choice partly restored by the paternalistic rule. Under-confident agents are, as we saw in a previous section, inclined to set unnecessary harsh personal rules and/or to sanction them with side-betting. Consider Stephanie, who is well aware of the risks generated by drinking too much. She is willing to reduce her consumption and to avoid the temptation of buying extra-large cups (despite them being cheaper). However, Stephanie – like E – does not trust herself with an 'Avoid buying extra-large cups of soda' rule and forces herself to adopt a 'Never buy sodas ever again' rule. When the law forbidding selling extra-large cups is enforced, she can dismiss her personal rule, and enjoy buying sodas again. In very specific cases, seemingly coercive legal rules can foster – rather than infringe – individual freedom of choice. Paternalistic strategies are not necessarily more coercive than their non-paternalistic alternatives.

I have shown so far that infringement of individual freedom is not a general feature of paternalism, and that, contrary to some common belief, paternalism is not necessarily more coercive than self-command and even can, in some cases, be a better alternative. I have shown, in particular, that much actually depends on the agents' level of sophistication and her degree of self-confidence. I would like now to elaborate on the relationship between legal paternalism and nudges, on one hand, and legal paternalism and personal rules, on the other hand. My aim here is to show, based on the results found so far, that it is misleading to oppose these different policies, as they usually work best together.

Table 2 shows that nudges are only efficient in intermediate cases, while legal paternalism is most beneficial in more extreme cases. Take, once more, the example of the cafeteria and suppose that customers have an initial preference for healthy food. Rearranging the dishes will have no effect on A, D or E. There is no point in nudging

A as she does not even have a personal rule to follow. As for D and E, the rule they imposed on themselves is sufficiently harsh (and even excessively harsh in the case of E) for them to prefer the healthy food whatever the dish arrangement is. B and C, on the other hand, can be affected in their choice by the line rearrangement. Because her willpower is only slightly overestimated, C is more likely to be nudged than B. So out of the five cases presented in Table 2, only C (and possibly B) can potentially be nudged [3]. On the other hand, A and E can only benefit from rigid and seemingly coercive rules. In the case of A, preference reversals are unavoidable even in the presence of nudges. Personal rules and self-commitment are also made impossible by the agent's extreme self-confidence. In a case like this, the only possible way to remedy DIBs (if this is justified) is to impose rigid rules. In the case of E, as explained above, laws are not only beneficial to solve time inconsistencies but they also have a negative cost compared to personal rules. A first conclusion is then that, contrary to what Sunstein and Thaler claim, nudges cannot be generalized and they certainly cannot be an alternative to legal paternalism. At best, nudges help slightly overconfident people.

A second important point is that personal rules and external ruling (or nudging) are complementary rather than alternative policies. As explained above, the relative cost of paternalism (in terms of opportunities of choice) functionally depends on the gap existing between the agent's actual and perceived willpower. Ideally, paternalistic laws should be modeled on the personal rules perfectly sophisticated agents – like D – give themselves. The more people deceive themselves the more costly the law becomes. In other words, the coerciveness of the law is the price of self-deception. But laws do not only sanction self-deception, they also contribute to improving people's sophistication. Paternalistic laws, when they are properly set and when all concerned know that they represent the personal rule that they themselves should have, give each agent the possibility to infer from it the extent of her self-deception. If all individuals were fully rational, they would use paternalistic laws to revise their level of self-confidence, until their personal rules and the paternalistic laws corresponded perfectly. Whether they actually are rational or not is another question.

To finish, I want to emphasize the limitations of this study. As I made clear at the beginning, my aim here is not to defend paternalism but to consider – as objectively as possible – the cost of paternalism relative to self-command. I conclude that paternalistic actions are not necessarily more costly than self-commanding ones, and

that legal paternalism can help overconfident agents to establish better personal rules. I have based my conclusions on a single example, supposing a common willpower, and different levels of self-confidence. The degree of willpower chosen  $\frac{2}{5}$  is voluntarily low to make the case more compelling. Likewise, I implicitly supposed that the case in question was serious enough to possibly prompt legal action. Yet I certainly do not believe that all instances of DIBs call for paternalistic actions (legal or not). To emphasize, my only purpose here was to assess the relative cost of paternalistic actions compared to non-paternalistic ones *when an action is needed*. It is clear that all time inconsistencies (since I specifically focused on those) do not call for an action and that some of them can possibly call for a personal action but not for a paternalistic one. It would certainly be helpful to know when paternalistic actions are socially tolerable before assessing them. Ultimately, I leave this question open for future debate.

## Conclusion

I claimed in this paper that the cost of paternalism, generally speaking, should not be assessed in absolute terms (in terms of opportunities of choice, for instance) but *relatively* to the cost of self-commanding strategies. Nudges may not be very constraining but they can be relatively more constraining than personal rules. Conversely legal rules can be coercive but comparatively less so than personal rules. If freedom of choice is to be adopted as moral criterion, paternalist policies should then be tested with comparative benefit-cost analysis. Comprehensive studies would undoubtedly show that legal paternalism is sometimes to be preferred to any other alternatives. There are different types of responses to irrational self-regarding actions, of which self-command, behavioral paternalism, legal paternalism, empathic paternalism (a type of paternalism I unfortunately had no room to investigate here) etc. None are absolutely good or better than the others, including self-commanding strategies which are highly regarded by liberal thinkers, but they all have a role to play. I argued that Sunstein and Thaler's paternalism is useful in specific cases but that it cannot be a substitute for legal paternalism or for personal rules. Nudges can improve the efficiency of personal rules but they are pointless in their absence. Legal rules, on the contrary, can (and ought to) stand in for personal rules when they are inexistent or when they are inadequate. In the great majority of cases, however, legal

and personal rules work together. Legal rules play an important role towards sophisticated people: they set a standard people can easily refer to.

## Endotes

[1] A second important distinction consists in telling apart the respective reactions of sophisticates and naïfs before immediate rewards and immediate costs. O'Donoghue and Rabin (1999) show that it is wrong to assume that naïfs are always worse off than sophisticates. If sophisticates are better equipped to fight off procrastination, for instance, they are prone to go too far the other way, preproperation. In a number of situations, such as this one, sophisticates' attempting to outbalance their time inconsistency can result in excessive prudence, or even negative time discounting. Sophistication is generally an advantage when it comes to time inconsistency, but self-medicated policies can have a substantial cost that ought to be included when assessing the relative (dis)utility of paternalism.

[2] One could argue that individuals can 'forget' past lapses and that this would therefore not affect their reputational capital. It is difficult to know exactly to what extent these memory lapses are conscious (bad faith) or unconscious. In any case they are very common, and can significantly counter-balance or exacerbate the self-enforcing nature of personal rules. On one hand, unremembered achievements do not contribute to bolstering reputational capital and ignored failures do not jeopardize it. If 'memory gaps' were equally distributed between successes and failures, the trend observed above would not be disproved. Yet empirical studies have shown that human beings are more likely to forget their failures than their successes. A logical consequence of this is that weak-willed individuals do not get weaker, but that they merely remain weak, whilst strong-willed individuals still get stronger. The gap between weak and strong willed individuals is still deepening but at a slower rate. On the other hand, selective memory can extend to testing one's willpower. The difference with the previous point is slight but important. In the previous case, weak-willed individuals forgot that they failed the test. They are nonetheless still aware of their time inconsistency. In the present case, weak-willed individuals forget even that they put their willpower to the test. This implies that they are now denying their preference reversal. Like the fox in front of the sour grapes, they prefer rationalizing their failure

Salvat, Christophe (2015), 'Economics of paternalism: The hidden costs of self-commanding strategies', *The Journal of Philosophical Economics. Reflections on Economic and Social Issues*, IX: 1, 102 - 124

rather than admitting it or even forgetting it. Their self-confidence remains intact, and they have now nothing to envy in the strong-willed individuals.

[3] It does not, however, mean that opportunities to nudge are relatively scarce. It could well be that 80% of the population belongs to categories B or C.

## References

- Ainslie, George (1992), *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person*, Cambridge, UK: Cambridge University Press.
- Akerlof, George A. (1991), 'Procrastination and Obedience', *The American Economic Review*, 81 (2), 1-19.
- Becker, Gary S. and Kevin M. Murphy (1988), 'A Theory of Rational Addiction', *Journal of Political Economy*, 96 (4), 675-700.
- Bénabou, R. and J. Tirole (2002), 'Self-Confidence and Personal Motivation', *The Quarterly Journal of Economics*, 117 (3), 871-915.
- Bénabou, R. and J. Tirole (2004), 'Willpower and Personal Rules', *Journal of Political Economy*, 112 (4), 848-886.
- Bovens L. (2009), 'The ethics of nudge' in Till Grüne-Yanoff and Sven Ove Hansson (eds), *Preference Change. Approaches from Philosophy, Economics and Psychology*, Netherlands : Springer, pp. 207-219.
- Christensen-Szalanski, JJ. (1984), 'Discount Functions and the Measurement of Patients Values', *Medical Decision Making*, 4, 47-58.
- Dworkin, G. (1983), 'Paternalism' in R. Sartorius (ed), *Paternalism*, Minneapolis: University of Minnesota Press, pp.19-34.



Salvat, Christophe (2015), 'Economics of paternalism: The hidden costs of self-commanding strategies', *The Journal of Philosophical Economics. Reflections on Economic and Social Issues*, IX: 1, 102 - 124

Elster, J. (1979), *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge, UK: Cambridge University Press.

Feinberg, J. (1986), *Harm to Self: The Moral Limits of the Criminal Law*, New York: Oxford University Press.

Goldman, SM. (1980), 'Consistent Plans', *The Review of Economic Studies*, 47 (3), 533-537.

Gul, F. and W. Pesendorfer (2001), 'Temptation and Self-Control', *Econometrica*, 69 (6), 1403-1435.

Gul, F. and W. Pesendorfer (2004), 'Self-Control and the Theory of Consumption', *Econometrica*, 72 (1), 119-158.

Gul, F. and W. Pesendorfer (2007), 'Harmful Addiction', *The Review of Economic Studies* 74 (1), 147-172.

Hedden, B. (2015) *Reasons without Persons: Rationality, Identity, and Time*, Oxford: Oxford University Press.

Herrnstein, R.J. and D. Prelec (1992), 'A Theory of Addiction' in G. Loewenstein and J. Elster (eds), *Choice over Time*, New York: Russell Sage Foundation, pp.331-360.

Holton, R. (2009), *Willing, Wanting, Waiting*, Oxford: Oxford University Press.

Laibson, D. (1997), 'Golden Eggs and Hyperbolic Discounting', *The Quarterly Journal of Economics*, 112 (2), 443-477.

Marneffe, P. de (2006), 'Avoiding Paternalism', *Philosophy and Public Affairs*, 34 (1), 68-94.

O'Donoghue, T. and M. Rabin (1999), 'Doing It Now or Later', *The American Economic Review*, 89 (1), 103-124.

Salvat, Christophe (2015), 'Economics of paternalism: The hidden costs of self-commanding strategies', *The Journal of Philosophical Economics. Reflections on Economic and Social Issues*, IX: 1, 102 - 124

O'Donoghue, T. and M. Rabin (2001), 'Choice and Procrastination', *The Quarterly Journal of Economics*, 116 (1), 121-160.

Peleg, B. and ME. Yaari (1973), 'On the Existence of a Consistent Course of Action when Tastes are Changing', *The Review of Economic Studies* 40 (3), 391-401.

Pollak, RA. (1968), 'Consistent Planning', *The Review of Economic Studies*, 35 (2), 201-208.

Rawls, J. (1971), *A Theory of Justice*, Cambridge, Massachusetts: Harvard University Press.

Salvat, C. (2014), 'Behavioral Paternalism', *Revue de Philosophie Economique/Review of Economic Philosophy*, 15 (2), 109-130.

Schelling, TC. (1960), *The strategy of conflict*, Cambridge, Massachusetts: Harvard University Press.

Schelling, TC. (1984), 'Self-Command in Practice, in Policy, and in a Theory of Rational Choice', *The American Economic Review*, 74 (2), 1-11.

Schelling, TC. (1992), 'Self-Command: A New Discipline', in G. Loewenstein and J. Elster (eds), *Choice over Time*, New York: Russell Sage Foundation, pp.167-176.

Strotz, RH. (1955), 'Myopia and Inconsistency in Dynamic Utility Maximization', *The Review of Economic Studies*, 23 (3), 165-180.

Christophe Salvat is researcher at the Centre National de la Recherche Scientifique (CNRS) and member of Triangle, UMR5206, at the Ecole Normale Supérieure de Lyon (christophe.salvat@ens-lyon.fr).



## BEHAVIORAL PATERNALISM

Christophe Salvat

Vrin | « Revue de philosophie économique »

2014/2 Vol. 15 | pages 109 à 130

ISSN 1376-0971

ISBN 9782711652112

Article disponible en ligne à l'adresse :

-----  
<https://www.cairn.info/revue-de-philosophie-economique-2014-2-page-109.htm>  
-----

Distribution électronique Cairn.info pour Vrin.

© Vrin. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

# BEHAVIORAL paternalism

CHRISTOPHE salvat\*

## Résumé

Un nouveau type de paternalisme s'est développé ces dix dernières années sous l'impulsion de travaux innovateurs de certains économistes comportementaux. Ce nouveau type de paternalisme, que j'appelle ici paternalisme comportemental, s'est popularisé grâce à la théorie du « coup de pouce » de Richard Thaler et Cass Sunstein et remet en question l'idée selon laquelle le paternalisme serait inacceptable dans nos sociétés. L'objet de cet article est d'évaluer sa légitimité morale sans, néanmoins, se limiter à son supposé libertarianisme. Les résultats de mon investigation peuvent se résumer ainsi : bien que le paternalisme comportemental soit généralement reconnu pour son caractère libéral, il ne satisfait pas en fait les conditions de ce que Joel Feinberg nomme le « paternalisme mou ». Néanmoins, il possède des qualités morales sous-estimées par ses partisans. Il résiste d'abord très bien à la critique égalitariste d'Elizabeth Anderson. A la différence des formes traditionnelles de paternalisme, le paternalisme comportemental n'est pas dégradant et n'est pas ostracisant. Le paternalisme comportemental, enfin, peut se targuer d'être véritablement altruiste, à la condition, cependant, d'abandonner les hypothèses principales de Sunstein et Thaler.

---

\* Christophe SALVAT est Chercheur CNRS, Triangle, UMR 5206, ENS de Lyon. Email : Christophe.salvat@ens-lyon.fr. Early versions of this paper have been presented at the Collège International de Philosophie in Paris (January 2010), at the University of Lille (Institut d'Economie Politique, June 2012), at the University of Strasbourg (BETA Seminar, November 2012) and at the University of Cambridge (Faculty of Philosophy, January 2013). I am grateful to all participants for their insightful remarks. I would also like to thank the referees of *Revue de Philosophie Économique* for their helpful comments.

**Mots-clés :** paternalisme, comportementalisme, éthique, égalité, autonomie

**Abstract**

Over the last decade a new type of paternalism has emerged thanks to the groundbreaking works of some behavioral economists. This new type of paternalism, that I call here Behavioral Paternalism (BP), has become popular through Sunstein and Thaler's Nudges theory and challenges the view that paternalism is unacceptable today. The aim of this paper is to assess its moral legitimacy (not exclusively focusing on its alleged libertarian nature). The results of my investigation can be summarized as follows. Though BP is usually acknowledged for its "libertarian" character, it does not satisfy the conditions of what is considered, since Feinberg, as "soft paternalism". Nevertheless, BP has a strong point that has been underestimated by its partisans: it first withstands Anderson's equality argument. Unlike traditional forms of paternalism, BP is not demeaning and does not ostracize any category of people. Lastly, BP can be proved genuinely altruistic. This, however, demands that one abandons Sunstein and Thaler's main assumptions.

**Keywords:** paternalism, behaviorism, ethics, equality, autonomy

**Classification JEL :** A12, D64, I18, I31, K32

## 1. INTRODUCTION

The idea of paternalism has suffered dramatically since the decline of the nineteenth-century model of entrepreneurial capitalism. As a social practice, it is widely considered as a patronizing and morally inappropriate conduct to adopt for any individuals benefiting from a superior economic, social or political position. The philosophical debate over the legitimacy of paternalism, however, twice endeavored to challenge this view. A first wave of arguments in favor of paternalism dates from the 1970s<sup>1</sup>, a second only started a few years ago under the influence of major behaviorist thinkers. It emerges from all these contributions that modern paternalism no longer bears resemblance to social or

---

1. Most of the original contributions as well as a number of unpublished ones (including Dworkin's *Second Thoughts*) have been gathered in the volume edited by Rolf Sartorius in 1983.

religious charity. In the 1970s paternalism was essentially considered from a legal point of view (Feinberg 1983, 1986; Dworkin 1971) or, if not, from a medical point of view (Gert and Culver 1976; Buchanan 1978). The questioning, however, remains the same: under what conditions can paternalism be morally acceptable?

In the nineteenth-century entrepreneurial or social paternalism was often perceived as serving the interests of the dominant class under false pretenses. Paternalism was thought immoral because it was not genuinely altruistic. Since WWII legal paternalism has been criticized for being intrusive and coercive. Paternalism was then thought immoral because it was perceived as a threat to individual freedom. A few years later, paternalism suffered another philosophical blow when it emerged that it was also violating individuals' basic right to equality (Dworkin 1978; Anderson 1999; Arneson 2005). As one can see, the moral demands made to paternalism have shifted over time and conformed to new ideologies. But none disappeared altogether. Theories defending paternalism have to therefore adjust to continually meet more demanding moral standards. It is reasonable to say that, paternalism, today, is considered illegitimate a) when it acts under false pretenses to satisfy the interest of the paternalist agent, b) when it violates the individual autonomy of the people interfered with and c) when it does not respect the equality between people by singling out those who are deemed unable to decide for themselves. I shall refer to these three moral requirements respectively as the altruistic, the liberal and the egalitarian provisos.

At the beginning of the new millennium, spurred on by the success of behavioral studies, new theories on paternalism took up the challenge (Jolls, Sunstein, and Thaler 1998; Thaler and Sunstein 2003, 2008; Camerer et al. 2003; Loewenstein and Haisley 2008). Sometimes called "libertarian", "asymmetric", "light" or simply "new" paternalism, these theories are all related to behavioral sciences. For simplicity's sake, I shall refer to them as Behavioral Paternalism (BP). Is BP really different from classic forms of paternalism? Does it really meet the moral criticism paternalism is facing? This is the object of this study. This paper does not pretend to be a defense (or a criticism) of BP, it is rather an acknowledgement of its theoretical potentialities, including those they have chosen not to pursue.

This paper is divided into four sections. In the first section I present this new theory of paternalism I call Behavioral Paternalism (Section 2). I demonstrate that behavioral theories of paternalism indeed differ from traditional forms of paternalism. The three following sections examine its moral robustness, which can be gathered under three moral provisos: the liberal proviso, the egalitarian proviso and the altruistic proviso. The issue of personal autonomy has been the most discussed so far. It has been argued that BP was the only form of paternalism capable of meeting the liberal proviso. I shall argue that this is not the case (Section 3). I believe, nevertheless, that BP definitively has a moral advantage over other types of paternalism, in its unique ability to meet Anderson's egalitarian critic (Section 4). It could also meet certain arguments generally made against the morally deceiving nature of paternalism, even though this is clearly not the path chosen by its defendants (Section 5).

## 2. BEHAVIORAL PATERNALISM

In this section I present a new approach to paternalism that is referred to sometimes as “New Paternalism” (Rizzo 2009; Mead 1997; Holt 2006; Economist 2006), “Asymmetric Paternalism” (Camerer *et al.* 2003; Loewenstein 2007 *et al.*), ‘Light’ (Loewenstein and Haisley 2008), “Positive” (Blumenthal 2007) or “Libertarian Paternalism” (Sunstein and Thaler 2003, 2008). Despite this variety of designations I nevertheless prefer to call it “Behavioral Paternalism” (hereafter BP). My justification for this is that, apart from “New Paternalism” (but being new is not a definitive characteristic), all these expressions prejudice the legitimacy of this new type of paternalism. Since my aim is precisely to discuss the terms of its legitimacy, I therefore believe that a neutral but informative qualifying adjective is more appropriate. The term “behavioral” imposes itself since the revival of paternalism observed for the last decade originates from behavioral sciences.

The American lawyer Cass Sunstein and economist Richard Thaler are the first and most important champions of BP. Their groundbreaking works on libertarian paternalism have produced a vast academic literature and found a favorable echo with politicians.



Their work, directly inspired by Richard Thaler's own contributions to behavioral economics, has the merit of raising anew the question of the possibility and the legitimacy of paternalism in the XXI<sup>st</sup> century. The paternalism they defend radically differs from the traditional philosophical approaches inherited from Mill's "soft anti-paternalism" (Feinberg 1986, p.15). They contest, in particular, the idea that paternalism is necessarily coercive and that it is predominantly institutionally-based.

Behavioral studies gave partisans of paternalism two major arguments. Firstly empirical studies have repeatedly shown that individual actions do not meet the rationality standard that the economic theory would expect. The idea of bounded rationality, introduced by Herbert Simon, has been decisive in explaining what was otherwise considered as economically inconsistent. For some years now, economic models of bounded rationality have been supplemented by new evidence provided by neuroscientists, giving birth to a new discipline called neuroeconomics (Camerer, Loewenstein, and Prelec 2004; Camerer, Loewenstein, and Prelec 2005; Camerer 2007; Kahneman 2011). According to Colin Camerer neurosciences definitely undermine the "Revealed-Preference Approach" by showing the role played by the difference between experienced utility ("liking") and decision-utility ("wanting") in individual suboptimal choices. Benevolent agents (or governments), who are aware of the gap between the utility individuals think they will obtain (what they want) and the utility (or disutility) they will actually get (what they like), are then given the opportunity to act paternalistically without risks (Camerer 2006, 2008). In his terms "paternalism could be justified, in terms of a person's own welfare, if the wanting system does not produce what the liking system likes and if the intervention creates more liking than the person would achieve on his own or with market-supplied help" (Camerer 2006, p.101-102).

A second argument for paternalism is individuals' endogenous preferences. Preferences are claimed by behaviorists to be endogenous to social contexts and consequently inconsistent from one situation to another, the lack of any past or similar reference

introducing an element of arbitrariness in the evaluation<sup>2</sup>. Over the last twenty years, behavioral economics has offered a great deal of evidence to support this thesis (Lichtenstein and Slovic 2006). Framing, anchor or endowment effects have been identified as some of the key mechanisms at stake (Thaler 1980, Loewenstein and Adler 1995; Tversky and Kahneman 1981, 1986). But if endogenous preferences are responsible for individuals' poor rational behavior, they can also be the medium for their own remedy. For Sunstein and Thaler, for instance, it is possible to significantly reduce the lack of individual rationality by switching the default rule settings from an "opt in" to an "opt out" formula. They illustrate the benefits of setting change in default rules with two case studies: the case of 401 (k) employee savings plans (Madrian and Shea 2001, Choi *et al.* 2002) and their own experience of the University of Chicago's car park payment system.

In the first case, employers decide that instead of letting their employees join a savings plan, the American 401 (k) plan, they would automatically enroll them unless employees specifically stated that they did not wish to be included. When eligible, employees usually receive a 401 (k) plan form that they must complete in order to join. In this case, however, employees only receive a statement in which they are told that, unless they opt-out, they are automatically enrolled in the plan. In both studies, a significant increase in the enrollment rate was observed. In the second case, a change in the tax law made it possible for employees to pay for employer-provided parking on a pre-tax basis. The University of Chicago presumed that, although being in the employees' interest, many of them would not take the trouble to fill in and send back the form and decided to enroll all employees automatically. Those who preferred to pay with after-tax dollars were, however, free to opt out<sup>3</sup>. Had the university chosen a different default rule (like an "opt

---

2. The idea of "coherent arbitrariness" of preferences developed by Dan Ariely, George Loewenstein and Drazen Prelec shows, however, that valuations are not entirely arbitrary: absolute valuation of goods or experience is for a large part arbitrary, but relative valuations can prove to be coherent (Ariely, Loewenstein and Prelec 2003).

3. In practice, however, default rules are sticky and guaranteeing the conditions of free choice is difficult. Once a person has been enrolled in a default rule, he or she tends to overestimate its benefits (endowment effect). If opting in is a deterrent for many people,

in” strategy), the authors reckon that “many employees, especially faculty members (and probably including the present authors), would still have that form buried somewhere in their offices and would be paying substantially more for parking on an after-tax basis” (Thaler and Sunstein 2003, p.1171). A minor effort, such as filling in a form, is often sufficient not to claim a benefit people are entitled to.

In the introduction I have suggested that – in order to be moral – paternalism ought to fulfill three criteria: respecting personal autonomy, respecting individual equality and being genuinely morally motivated. Let us now consider these three criteria in connection with behavioral paternalism.

### 3. THE LIBERAL PROVISIO

The first and most intuitive argument opposed to paternalism is the respect of individual freedom. It is by far the one that has been given the most attention. Opponents of paternalism defend a right to autonomy or self-direction. It is naturally not an absolute right. The nature of governments is to interfere with an individual’s decision-making process. Liberals believe that it is legitimate to do so only when individual actions are other-regarding. They hence justify legal restrictions such as the obligation for an individual to wear a helmet or a seatbelt whilst driving a car or a motorbike by the benefit reaped by others (the cost borne by the collectivity) rather than by the utility the individual should expect from the measure. This means that to comply with this critic, self-regarding justifications for paternalism must always be proved inadequate to back interventions or they must be able to be otherwise justified. In practice, however, it is rather difficult to ascertain that no self-regarding justification for paternalism can ever be found or that other-regarding justifications will be sufficient. The question is

---

opting out from a default rule is even more dissuasive. To respect freedom of choice, a libertarian planner must set up a default rule that will be easy and costless for individuals to opt out of. Ideally, it should not be more demanding than a ‘one-click’ procedure (Thaler and Sunstein 2008, p. 249).

therefore not whether paternalism is a legitimate practice but to what extent autonomy provides a justification (or not) for paternalism.

To remedy the totalitarian tendency of paternalism, Joel Feinberg proposes to vet its practice on the condition that it respects individual freedom. This “soft” or “weak” version of paternalism originates from Mill’s own ambiguity in his famous harm principle:

That principle is, that the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection. That the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. [...] He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because in the opinion of others, to do so would be wise, or even right. (Mill 1989, p.13)

Due to the ambiguity of Mill’s phrasing (“against his will”) – but also due to a pervasive ambiguity throughout his work (Arneson 1980; Claeys 2013; Brink 2013) – the discussion about the legitimacy of paternalism soon boiled down to a single argument: the degree of voluntariness in the action interfered with. The notion of will can indeed either refer to a current (and temporary) desire or to a “true” and rational will. From this information the nature (and hence the legitimacy) of paternalism could be inferred: a weak or soft paternalism meant to protect individuals against self-inflicting (but not fully voluntary) harm, while “in its extreme version”, a strong or hard paternalism was supposed “to guide them, whether they like it or not, toward their own good” (Feinberg 1983, p.3).

For Joel Feinberg, the violation of individual autonomy is the sole possible justification for anti-paternalism. Conversely autonomy can also be its sole possible justification. In a certain number of cases paternalistic interventions can thus be justified by the absence or insufficiency of individual autonomy. Four factors contribute, for Feinberg, to weaken the autonomous or voluntary quality of individual decisions. Those are:

### 1. Coercion or threat of coercion

2. Psychological dependency (drug, alcohol, hypnosis...)
3. Emotional ascendency (depression, anger, obsession...)
- 4a. Misinformation, deceit or wrong belief that misleads the agent relative to the consequences of his action (i.e. accidentally taking arsenic for salt)
- 4b. Misinformation or a wrong belief (produced by the individual inability to gather correct information) that eventually misleads him relative to the consequences of his action.

Voluntariness is, for Feinberg, essentially based on responsibility. The five above circumstances obliterate the decision maker's responsibility since the choices made do not express the individual character. Conversely, individuals take full responsibility for autonomous or voluntary choices "since they represent him faithfully, expressing his settled values and preferences" (Feinberg 1989, p.113). For Feinberg, misinformation always constitutes an autonomy-diminishing factor<sup>4</sup>.

Soft or weak paternalism typically describes situations of involuntary acts or misinformed decisions that must be interfered with for the person's own good. Typically, a soft paternalist can legitimately prevent a person from using the saltcellar if they have not been informed that it contains arsenic instead of salt. Similarly, a person under the influence of drugs or alcohol can be temporarily prevented from making decisions that he/she may bitterly regret the next day. Alternatively, if a person is appropriately informed of the risks incurred from crossing a dangerous bridge, smoking, or dueling, then no external interference is ever justifiable to stop that person carrying out such an action. Soft paternalism re-establishes what some regard as an essential feature of liberalism: the right to act irrationally.

Hard paternalism, on the contrary, is by nature coercive. According to Gerald Dworkin, for instance, paternalism covers "the

---

4. Richard Arneson argues, however, that "a person's actions may be authentic expressions of his personality without being deliberately chosen" (Arneson 1980, p.487). 4b is thus only acceptable if the behavior responsible for the misinformation is itself 'out of character', that is, which is submitted to one of the previous autonomy-diminishing factors. He concludes that voluntary choice is admittedly "important but [that it] does not plausibly have the make-or-break significance that soft paternalism attaches to it" and that it is therefore "a mistake to make a fetish of" it (Arneson 2005).

interference with a person's liberty of action justified by reasons referring exclusively to the welfare, good, happiness, needs, interests or values of the person coerced" (Dworkin 1972, p.65). It legitimizes interference from voluntary and well-informed actions. Among all the possible examples, Dworkin mentions laws requiring motorcyclists to wear safety helmets, laws regulating sexual conduct, laws forbidding people from swimming at a public beach when lifeguards are absent and laws compelling people to save a fraction of their income for retirement. Although some of these examples would not be considered as paternalistic today, either because they involve a third element (the economic cost borne by the society) or because they are no longer considered self-destructive or wrong (homosexuality), these examples are representative of an essential feature of hard paternalism: the source of risk assessment. Whilst supporters of soft paternalism consider risk assessment as the risk taker's responsibility, risk is externally assessed by the paternalistic agent.

Defendants of BP trust that its main asset is not to be coercive. This is the reason why they described it as "libertarian", "soft", "light" or "asymmetric". Sunstein and Thaler describe their approach as "a relatively weak and nonintrusive type of paternalism" (Sunstein and Thaler 2003, p.1162). This, I believe, is grossly misleading. Sunstein and Thaler's libertarian paternalism borrows features from these two opposing approaches. It is not a form of soft paternalism. It is, on the contrary, rooted in a very weak notion of voluntariness that makes it closer to hard paternalism.

The libertarian feature of their theory relies exclusively on freedom of choice. Yet, their analysis, based on behavioral studies, shows that most individuals' choices are sub-optimal because their valuations depend on arbitrary anchors and frames. To a certain degree, most of their actions are then *involuntary*: they want to be thin but eat fatty foods, they want to have a comfortable retirement but don't save enough money, they want to donate their organs but don't say so etc. Their actions are involuntary because determinant information is hidden from individuals (like in Mill's dangerous bridge case) (4a) but also because they believe individuals are naturally inapt to correctly treat the information they have (anchors, framing effects) (4b).

There is definitely an incongruity at simultaneously defending absolute freedom of choice and a weak concept of voluntariness. By downsizing individual responsibility, the authors actually undermine the possibility of a liberal paternalism. BP, against all expectations, does not comply any more with individual freedom of choice than any other forms of paternalism. Partisans of paternalism would be misled to think they have found there an answer to the liberal critic. They have, however, other reasons to rejoice. Its main moral asset is not there. What BP does really change for paternalism is its relation to equality.

#### 4. THE EGALITARIAN PROVISIO

The second criticism addressed to paternalism is that it violates the fundamental principle of equality. Many philosophers understandably condemn paternalism for its condescending and offensive practices (Shiffrin 2000; Anderson 1999; Kleinig 1983; VanDeVeer 1986). A number of liberal thinkers consider the right to equality second to none (Dworkin 1978; Anderson 1999; Rawls 1971; Arneson 2000), but they often disagree on what constitutes an unfair inequality. Paternalism primarily originates from the desire to correct inequality of luck, that is to say inequalities that do not result from individual choices. There is, for instance, little merit in becoming deputy director of a company when one is born in a rich and well-educated family in the nineteenth century, especially when the company belongs to one's father. It is conversely extremely difficult to access such a position when one is born in a working class family or a poorhouse orphanage. This is what is called (bad) brute luck. This is the reason why, historically, the wealthy and the well-educated often felt the need to share what they believed to be undeserved luck with the working classes by providing them with free access to education. Although intuitively very appealing, luck egalitarianism has come under criticism for its demeaning attitude towards those who were deemed to be victims of bad luck. What Anderson reproaches about luck egalitarianism is that it categorizes the victims of bad luck according to what they are (stupid, untalented, disabled, ugly, socially awkward, etc.). She argues that the point of equality is neither resources nor luck but status. And

treating a category of people as children or adults as “too stupid to run their lives” (Anderson 1999, p.301) is incompatible with the respect all individuals should equally be granted. She writes

Egalitarianism ought to reflect a generous, humane, cosmopolitan vision of a society that recognizes individuals as equals in all their diversity. It should promote institutional arrangements that enable the diversity of people’s talents, aspirations, roles, and cultures to benefit everyone and to be recognized as mutually beneficial. Instead, the hybrid of capitalism and socialism envisioned by luck egalitarians reflects the mean-spirited, contemptuous, parochial vision of a society that represents human diversity hierarchically, moralistically contrasting the responsible and irresponsible, the innately superior and the innately inferior, the independent and the dependent. It offers no aid to those it labels irresponsible, and humiliating aid to those it labels innately inferior. (Anderson 1999, p.308)

No wonder that she finds its response, paternalism, offensive. Unsolicited, paternalist actions have disastrous effects on the ill-endowed people they are supposed to help. By granting people a (financial) compensation for their condition, they actually treat them as incomplete human beings. In its most traditional form (derived from the father-child analogy), paternalism is also often justified by the economic, cognitive or social superiority of the paternalist agent<sup>5</sup>.

I believe, however, that behavioral paternalism can partly elude these criticisms. It is true that paternalism, generally speaking, is usually motivated by a willingness to correct an (unfair) inequality between people. It thus implicitly establishes a distinction between the ‘lucky’ and “unlucky” ones. Amongst the “unlucky” ones, paternalism particularly favors those who have no personal responsibility in their misfortune (“brute bad luck”): the “disabled”, the “ugly”, the “stupid”, and the “socially awkward” ones (Anderson, 1999). But paternalism is only demeaning if it finds its justification in such categories. I claim, however, that this ought not to be the case, and

---

5. For Gert and Culver, for instance, “paternalism can be practiced by anyone who has qualifications which he believes enable him to see better than S what is good for S’s good” (Gert and Culver 1976, p. 50).



that a morally legitimate paternalism is possible if it is grounded on positional asymmetries.

This means, in particular, that a superior economic rationality (a higher ability to compute or to order preferences) does not qualify as a good reason to interfere with someone else's life. In addition to being demeaning for those deemed to be intellectually ill-endowed, it has disastrous totalitarian consequences as it logically grants the most intelligent individuals the ability to overturn the decisions of the normally endowed ones "for their own sake" (Wikler 1979). By contrast, it is – I contend – morally acceptable for a person to act when her relative position gives her a cognitive advantage. Take Mill's famous example of the threatening bridge. Consider that A only knows about the poor state of the bridge because he is a local, whereas B – who is about to cross it – is foreign to the area. If B were local himself, he would not need A's warning and would also probably warn foreigners against the danger of this bridge. A is in a relatively better position to make a decision on this matter than B. This situation does not, however, say anything about A being inherently superior to B.

The same argument can apply to BP. It is possible to show that in the case of BP, paternalist agents cannot claim any superiority of body or mind. Take the cafeteria example: the director, or whoever arranges the dishes, is no different from her customers. In fact, she is a customer herself on her days off. The customers of the cafeteria are not victims of brute luck, waiting for her help. And she has no reason whatsoever to be kind to them (Sunstein and Thaler wrongly claim that she is necessarily benevolent, see next section). Nothing distinguishes her from the "paternalized customers" except the position she occupies at the decision-making moment. If she does arrange the dishes so that her customers eat healthier, then her action undoubtedly qualifies as paternalistic. There is nothing demeaning about it: customers are not disgraced or marked out and the director is not motivated by a sense of justice. She arranges the dishes relatively to what she considers to be a "good" order: meaning good for her and good for the others. BP does not require a cognitive or moral superiority from the paternalist agents. BP is what could be referred to as an "opportunistic paternalism". If anything, paternalism reflects individuals' equality in front of behavioral bias. It is rather unfortunate that Sunstein and Thaler did

not appreciate this point in their works, as it could have been one of their strongest arguments. They preferred seeing paternalist agents as “planners” or “choice architects”, which admittedly damages their chance of defending their paternalistic theory on egalitarian grounds.

In the previous paragraphs I have demonstrated 1) that BP does not conform to the liberal proviso but 2) that it is the only kind of paternalism that actually comply with the equality proviso. The last point to be considered is its altruistic nature.

## 5. THE ALTRUISTIC PROVISIO

The third criticism addressed to paternalism is that under the false pretension of being altruistic, it actually only serves the interest of the paternalistic agent. Paternalism is traditionally accused of being immoral because it is self-interested and that it uses immoral means to conceal its real nature. The captains of industry of the nineteenth-century who built social houses, school and hospitals for their workers have often been accused of being more interested in the productivity and the low turnover of their company than by the actual well-being of their low paid and hardworking employees. Besides, it is not difficult to imagine cases where “good natured people” resort to “wrong means” in order to complete what they consider as being “a good action”<sup>6</sup>. There is undoubtedly some truth in this. There is no point in denying it. Paternalism has indeed suffered from a long history of abusive practices. But in that respect, I believe that paternalism does not significantly differ from

---

6. Allen Buchanan, for instance, has condemned a number of disgraceful medical practices involving withholding information or even bluntly lying to patients about their state of health to ‘avoid them any unnecessary harm’ (Buchanan 1978). Bernard Gert and Charles Culver, studying the moral practices of medicine (a field known today as bioethics), also condemn the shams of medical paternalism (Gert and Culver 1979). They contend in particular that “an essential feature of paternalistic behavior toward a person is the violation of moral rules (or doing that which will require such violations), for example, the moral rules prohibiting deception, deprivation of freedom or opportunity, or disabling” (Gert and Culver 1976, p. 48). Most of the criticisms addressed to medical paternalism can nevertheless be handled through the principle of equality put forward by Elizabeth Anderson.

benevolence: a large number of benevolent actions are also self-interested but it does not imply that benevolence, as a general practice, is wrong. So let us admit that moral actions do exist, and that paternalism can be genuinely altruistic. By altruistic I mean disinterested and benevolent. I do not contest the possibility that the agent might have some indirect and/or long-term interest in his benevolent action. After all, we all ultimately gain from altruism. I merely consider that his action is altruistic if it is not *primarily* motivated by self-interest. To borrow Thomas Nagel's phrase, it is "a willingness to act in consideration of the interests of other persons, without the need of ulterior motives" (Nagel 1970, p.79). Actions that are primarily motivated by self-interest and concealed as altruistic can be discarded as *falsely* paternalistic. The difficulty, so far, has been to prove the possibility of genuinely paternalistic action.

Consider Sunstein and Thaler's famous example of the cafeteria. The director is aware that the way she arranges the line of dishes will "frame" the customers' choice. According to the authors, she faces four alternatives (Sunstein and Thaler 2003, p.1164):

1. She could make choices that she thinks would be best for the customers, all things considered.
2. She could make choices at random.
3. She could choose those items that she thinks would make the customers as obese as possible.
4. She could give customers what she thinks they would choose on their own.

Suppose for the moment that she is well-aware that customers (much like herself) know what is healthy for them and what is not. Suppose, also, that she is also aware (as her customers are as well) that the colors, the shapes, the odors and the disposition of goods arranged in front of them can influence their choice towards the unhealthiest dishes. Being director of the cafeteria and being in charge of the disposition of the line of dishes, she therefore has the opportunity to rearrange the dishes in order to reduce this behavioral bias. So, how likely is it for the director of the cafeteria to choose the altruistic option? Sunstein and Thaler claim that she will necessarily make the "right" choice (Sunstein and Thaler 2003,

p.1164). For one thing, it is largely unsupported (Blumenthal-Barby 2013), and for another, it is morally unjustifiable.

Supporters of BP are indeed stuck in-between two different but equally damaging positions. The first one consists in *defaulting* to altruism rather than deliberately choosing it. Robert Sugden once pointed out that one option is surprisingly missing from the list: the cafeteria director might simply choose to display the dishes in the most economically efficient way (Sugden 2008). The authors' reasons for not including it were twofold. First they argued that planners are not necessarily looking for profit maximization. Even directors of private companies, like school cafeterias, are not constantly subject to market pressures (Sunstein and Thaler 2003, p.1165). Secondly, and more importantly, they claimed that

market success will come not from tracking people's *ex ante* preferences, but from providing goods and services that turn out, in practice, to promote their welfare, all things considered. Consumers might be surprised by what they end up liking; indeed, their preferences might change as a result of consumption. (Sunstein and Thaler 2003, p.1165)

This basically comes down to acknowledging the fact that paternalist agents indeed have ulterior self-interested motives, effectively ruining any endeavor to restore the morality of paternalism. Default choices are not genuinely disinterested. From a libertarian vantage point, moreover, defaulting is not an option. Self-ownership theoretically excludes the possibility of individuals not having the choice not to be benevolent<sup>7</sup>.

A second possibility would be for BP to claim that paternalist agents are morally unimpeachable, and that they deliberately choose to be paternalistic out of sheer altruism. Paternalist agents would

---

7. Sunstein and Thaler acknowledge this libertarian principle. They however argue that if benevolence cannot be forced, it can be artificially prompted while freedom of choice is respected. The different systems of organs donations adopted by European and American countries illustrate that possibility. When Americans need to actively express their consent to be removed organs when they die, some European countries such as France presuppose that French implicitly consent to give organs unless explicitly stated. The authors reckon that the French setting of the default rule does not infringe on individuals' freedom of choice and that a change in the American policy could save thousands of lives each year. They call it *libertarian benevolence*.

then be composed of morally righteous individuals having the opportunity to help the others by engineering their decision-making context. But this would not do either. Moral righteousness would undoubtedly violate the equality proviso described above. Paternalism has been widely (and partly justly) criticized for letting the self-proclaimed “uppermost intelligent” or “supremely rational” individuals interfere with those they considered “too stupid to run their lives”. But what if they were deemed morally “superior”? What if they owed their status to their moral rather than to their cognitive excellence? Paternalism would still be unjustifiable. At first sight, BP seems to fail to provide a reasonable solution to the altruistic proviso. In fact, it seems to be even less capable of defending itself against this kind of criticism than any other forms of paternalism. Supporters of traditional paternalism can claim that altruism does exist and that it is crucial to differentiate between genuine and false paternalistic practices.

There is a way out of this dilemma; a way offered by behavioral studies yet largely ignored by BP’s own supporters, except maybe for Jeremy Blumenthal (Blumenthal 2005, 2007, 2013). It consists in explaining altruistic actions through emotional biases<sup>8</sup>. Consider the

---

8. Behavioral biases divert decision-makers away from their “ideal” choice in morals as well as in prudence. Numerous experimental studies support this statement, a number of which have been presented by John Doris in a compelling book on individuals’ (absence of) character (Doris 2002). Amongst the vast array of experiments quoted by Doris, let me mention two cases to illustrate the importance of framing effects in moral choice. The first one, undertaken in 1972 by the American psychologists Alice Isen and Paula Levin, took place in a shopping mall near a phone box (Isen and Levin 1972). Passersby and customers using the phone box were the unwitting participants of the experiment. At the end of each call, a young woman “accidentally” dropped a folder, thus spreading all her papers across the floor. The aim of the experiment was to study whether individuals are more likely to behave altruistically when they are in a good mood. Phone callers are divided in two groups, a control group for which nothing is done and a test group composed of good-humored subjects. In order to put the subjects in a good mood the experimenters place a “forgotten” dime in the phone receiver for the callers to find. The results of the study were striking: 14 out of 16 callers who found the dime came to the aid of the ‘clumsy’ young woman, whilst only one out the 25 who did not find the dime did! The second study belongs to a series of experiments led by the social psychologist Robert A. Baron (Baron and Thomley 1994; Baron and Bronfen 1994; Baron 1997) highlighting the effect of aromas on human behavior. In the last experiment Baron shows that individuals located near to a pleasant source of aromas (such as hot croissants or coffee) were more disposed to help passersby (in need of change in his experiment) than those who were situated in a

cafeteria example above. Suppose now that the director has to order fifteen dishes equally distributed between three classes of profitability level. She has to order all dishes, including the equally profitable ones. Since her interest as manager of the cafeteria is already satisfied, she is otherwise indifferent to the way in which dishes are arranged within each profitability category. She could, for instance, adopt a random criterion of distribution<sup>9</sup>. Random choice set apart there are many other possible criteria to arrange dishes: one can order them by size, by color, by shape, by composition, by season, by name, by origin etc. Altruism (or conversely malevolence) is then just one of many options available to the director of the cafeteria, which she chooses *according to her mood* or to irrelevant circumstances. It is absolutely fundamental that she does not feel the need to systematically act altruistically and that she is not nudged herself to do so. Paternalism is not and must not be inevitable if it wants to be morally acceptable.

The fact that individuals who have the opportunity to act paternalistically do not systematically decide to do so has a second advantage: it limits the practice of paternalism and consequently contributes to its social acceptability. Most of the philosophical debate on freedom focused on the legitimacy of paternalistic actions taken separately. Very little attention has been given, however, to the disagreement and the sense of violation felt by individuals who continuously have their choices interfered with, however legitimate each of these interferences might be. All paternalistic opportunities should not be exercised, however legitimate or well-intentioned they are. If everyone were to act altruistically our life would soon become absolutely unbearable.

---

neutral aromatic environment (Baron 1997). All these observations, as well as many others, corroborate recent neurobiological studies that show the impact of (positive) emotions on altruism (Damasio 2004).

9. Can a choice be absolutely random? Statisticians and neuroscientists have observed that the human brain is ill-disposed toward randomness and that individuals frequently prefer adopting any given criterion of choice rather than (unsuccessfully) attempting to simulate a random statistical distribution (Mlodinow 2009).

## 6. CONCLUSION

In this article I discussed whether the new type of paternalism defended by Sunstein and Thaler (amongst others) is better prepared to respond to moral attacks than any other type of paternalism so far. Contrary to a common view, behavioral paternalism is not better equipped to meet liberal attacks made against paternalism since John Stuart Mill. I demonstrated that it still does infringe on individual freedom and it certainly cannot be described as “libertarian”. To some extent, it is a much harder form of paternalism than Feinberg’s legal paternalism. It does not succeed either in answering to Anderson’s devastating arguments against the condescending nature of paternalism, or even to prove the fact that genuine disinterested paternalistic actions could exist. In that respect, behavioral paternalism fails to make a case for paternalism. Yet there is something in this class of theories that is definitely worth thinking about, despite its current theoretical shortcomings. Behavioral studies do bring something new to the field, something that well could be further developed as a proper defense of paternalism. For some reason, supporters of BP ignore these potentialities to dredge up the old debate on paternalism and freedom, ultimately showing that they do not consider BP as a completely new approach to paternalism. This would have been, however, their best shot.

## REFERENCES

- ANDERSON Elizabeth. 1999. “What Is the Point of Equality?”. *Ethics*, vol.109 n°2, pp.287-337.
- ARNESON Richard. 1980. “Mill versus Paternalism”. *Ethics*, vol.90 n°4, pp.470-489.
- . 2000. “Luck Egalitarianism and Prioritarianism”. *Ethics*, vol.110 n°2, pp.339-349.
- . 2005. “Joel Feinberg and the Justification of Hard Paternalism”. *Legal Theory*, n°11, pp.259-84.
- ARIELY Dan, George Loewenstein and Drazen PRELEC. 2003. “Coherent Arbitrariness: Stable Demand Curves without Stable Preferences”. *The Quarterly Journal of Economics*, n°118, pp.73-105.
- BARON Robert A. 1997. “The Sweet Smell of... Helping: Effects of Pleasant Ambient Fragrance on Prosocial Behavior in Shopping Malls”. *Personality and Social Psychology Bulletin*, vol.23 n°5, pp.498-503.

- BARON Robert A., Marna I. BRONFEN. 1994. "A Whiff of Reality: Empirical Evidence Concerning the Effects of Pleasant Fragrances on Work-Related Behavior". *Journal of Applied Social Psychology*, vol.24 n°13, pp.1179-1203.
- BARON Robert A., Jill THOMLEY. 1994. "A Whiff of Reality: Positive Affect as a Potential Mediator of the Effects of Pleasant Fragrances on Task Performance and Helping". *Environment and Behavior*, vol.26 n°6, pp.766-784.
- BLUMENTHAL-BARBY J.S. 2013. "Choice Architecture: A Mechanism for Improving Decisions while Preserving Liberty?". In C. Coons and M. Weber (eds.), *Paternalism. Theory and Practice*, pp.178-196. Cambridge: Cambridge University Press.
- BLUMENTHAL J.A. 2005. "Does Mood Influence Moral Judgment?: An Empirical Test with Legal and Policy Implications". *Law and Psychology Review*, n°29, pp.1-28.
- . 2007. "Emotional Paternalism". *Florida State University Law Review*, n°35, pp.1-72.
- . 2013. "A Psychological Defense of Paternalism". In C. Coons and M. Weber (eds.), *Paternalism. Theory and Practice*, pp.197-215. Cambridge: Cambridge University Press.
- BRINK David. 2013. *Mill's Progressive Principles*. Oxford: Oxford University Press.
- BUCHANAN Allen E. 1978. "Medical Paternalism". *Philosophy and Public Affairs*, vol.7 n°4, pp.371-390.
- CAMERER Colin. 2006. "Wanting, Liking, and Learning: Neuroscience and Paternalism". *The University of Chicago Law Review*, vol.73 n°87, pp.87-110.
- . 2007. "Neuroeconomics: Using Neuroscience to Make Economic Predictions". *The Economic Journal*, vol. 117 n°519, pp.26-42.
- . 2008. "The Case of Mindful Economics". In A. Caplin and A. Schotter (eds.), *The Foundations of Positive and Normative Economics: A Handbook: A Handbook*, pp. 43-69. New York: Oxford University Press.
- CAMERER Colin, Samuel ISSACHAROFF, George LOEWENSTEIN, Ted O'DONOGHUE, and Matthew RABIN. 2003. "Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism'". *University of Pennsylvania Law Review*, vol.151 n°3, pp.1211-1254.
- CAMERER Colin, George LOEWENSTEIN, and Drazen PRELEC. 2004. "Neuroeconomics: Why Economics Needs Brains". *Scandinavian Journal of Economics*, vol.106 n°3, pp.555-579.
- . 2005. "Neuroeconomics: How Neuroscience Can Inform Economics". *Journal of Economic Literature*, vol.43 n°1, pp.9-64.
- CHOI James, David LAIBSON, Brigitte MADRIAN, and Andrew METRICK. 2002. "Defined Contribution Pensions: Plan Rules, Participant Decisions, and the Path of Least Resistance". *Tax Policy and the Economy*, n°16, pp.67-114.
- CLAEYS Gregory. 2013. *Mill and Paternalism*. Cambridge: Cambridge University Press.
- DAMASIO Antonio. 2004. *Looking for Spinoza. Joy, Sorrow and the Feeling Brain*. London: Vintage.
- DORIS John M. 2002. *Lack of Character. Personality and Moral Behavior*. New York: Cambridge University Press.
- DWORKIN Gerald. 1971. "Paternalism". In Joel Feinberg and Hyman Gross *Philosophy of Law*, pp. 209-219. Belmont: Wadsworth Publishing Company.
- . 1972. "Paternalism". *The Monist*, n°1, pp.64-84.
- DWORKIN R.M. 1978. *Taking Rights Seriously*. Harvard: Harvard University Press.
- Economist (The)*. 2006. "The new paternalism". *The Economist*, 6th April.



- FEINBERG Joel. 1983. "Legal Paternalism". In Rolf Sartorius (ed.), *Paternalism*, pp.3-18. Minneapolis: University of Minnesota Press.
- . 1986. *Harm to Self: The Moral Limits of the Criminal Law*. New York: Oxford University Press.
- . 1989. *The Moral Limits of the Criminal Law: Volume 3: Harm to Self*. New York: Oxford University Press.
- GERT Bernard and Charles M. CULVER. 1976. "Paternalistic Behavior". *Philosophy and Public Affairs*, vol.6 n°1, pp.45-57.
- . 1979. "The Justification of Paternalism". *Ethics*, vol.89 n°2, pp.199-210.
- HOLT Jim. 2006. "The New, Soft Paternalism". *The New York Times*, 3rd December.
- ISEN A.M. and P.F. LEVIN. 1972. "Effect of Feeling Good on Helping: Cookies and Kindness". *Journal of Personality and Social Psychology*, n°21, pp.384-388.
- JOLLS Christine, Cass R. SUNSTEIN, and Richard THALER. 1998. "A Behavioral Approach to Law and Economics". *Stanford Law Review*, vol.50 n°5, pp.1471-1550.
- KAHNEMAN Daniel. 2011. *Thinking, fast and slow*. London: Allen Lane.
- KLEINIG J. 1983. *Paternalism*. Manchester: Manchester University Press.
- LICHTENSTEIN S., and P. SLOVIC. 2006. *The Construction of Preference*. Cambridge: Cambridge University Press.
- LOEWENSTEIN George and Daniel ADLER. 1995. "A Bias in the Prediction of Tastes". *The Economic Journal*, vol.105 n°431, pp.929-937.
- LOEWENSTEIN G., T. BRENNAN, K. G. VOLPP. 2007. "A Symmetric paternalism to improve health behaviors". *Journal of the American Medical Association*, vol.298 n°20, pp.2415-2417.
- LOEWENSTEIN George, F. and Emily HAISLEY. 2008. "The Economist as 'Therapist: Methodological Ramifications of 'Light' Paternalism". In Andrew Caplin and Andrew Schotter (eds.), *The Foundations of Positive and Normative Economics. A Handbook*, pp. 210-245. New York: Oxford University Press.
- MADRIAN Brigitte C. and Dennis F. SHEA. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior". *The Quarterly Journal of Economics*, vol.116 n°4, pp.1149-1187.
- MEAD L.M. 1997. *The New Paternalism: Supervisory Approaches to Poverty*. Washington: Brookings Inst Press.
- MILL John Stuart. 1989. *On Liberty and other writings*. Edited by Raymond Geuss and Quentin Skinner, *Cambridge Texts in the History of Political Thought*. Cambridge: Cambridge University Press.
- MLODINOW Leonard. 2009. *The Drunkard's Walk: How Randomness Rules Our Lives*. London: Penguin Books.
- NAGEL Thomas. 1970. *The Possibility of Altruism*. Princeton: Princeton University Press.
- RAWLS John. 1971. *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press.
- RIZZO Mario J. 2009. "Little Brother Is Watching You: New Paternalism on the Slippery Slopes". *Arizona Law Review*, vol.51 n°3, pp.685-739.
- SARTORIUS R.E. 1983. *Paternalism*. Minneapolis: University of Minnesota Press.
- SHIFFRIN Seana Valentine. 2000. "Paternalism, Unconscionability Doctrine, and Accomodation". *Philosophy & Public Affairs*, vol.29 n°3, pp.205-250.
- SUGDEN Robert. 2008. "Why incoherent preferences do not justify paternalism". *Constitutional Political Economy*, vol.19 n°3, pp.226-248.

- SUNSTEIN CASS R. and Richard H. THALER. 2003. "Libertarian Paternalism Is Not an Oxymoron". *The University of Chicago Law Review*, vol.70 n°4, pp.1159-1202.
- THALER Richard H. 1980. "Toward a positive theory of consumer choice". *Journal of Economic Behavior & Organization*, vol.1 n°1, pp.39-60.
- THALER Richard H. and Cass R. SUNSTEIN. 2003. "Libertarian Paternalism". *The American Economic Review*, vol.93 n°2, pp.175-179.
- . 2008. *Nudge. Improving Decisions About Health, Wealth, and Happiness*. New Haven & London: Yale University Press.
- TVERSKY Amos and Daniel KAHNEMAN. 1981. "The Framing of Decisions and the Psychology of Choice". *Science*, vol.211 n°4481, pp.453-458.
- . 1986. "Rational Choice and the Framing of Decisions". *The Journal of Business*, vol.59 n°4, pp.S251-S278.
- VAN DEVEER D. 1986. *Paternalistic Intervention: The Moral Bounds of Benevolence*. Princeton: Princeton University Press.
- WIKLER Daniel. 1979. "Paternalism and the Mildly Retarded". *Philosophy and Public Affairs*, vol.8 n°4.





**Astérior**

Philosophie, histoire des idées, pensée politique

12 | 2014

**Le principe de la folie et de la raison. Association des idées et liaison des idées aux XVII<sup>e</sup> et XVIII<sup>e</sup> siècles**

---

## Rousseau et la « Renaissance classique » française (1898-1933)

Christophe Salvat

---



### Édition électronique

URL : <http://journals.openedition.org/asterion/2545>

DOI : [10.4000/asterion.2545](https://doi.org/10.4000/asterion.2545)

ISSN : 1762-6110

### Éditeur

ENS Éditions

Ce document vous est offert par Aix-Marseille Université (AMU)



### Référence électronique

Christophe Salvat, « Rousseau et la « Renaissance classique » française (1898-1933) », *Astérior* [En ligne], 12 | 2014, mis en ligne le 24 juin 2014, consulté le 12 juin 2020. URL : <http://journals.openedition.org/asterion/2545> ; DOI : <https://doi.org/10.4000/asterion.2545>

---



Astérior est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.



## Rousseau et la « Renaissance classique » française (1898-1933)

### Christophe Salvat

Christophe Salvat est chargé de recherches au CNRS. Il est membre de Triangle (UMR 5206, ENS de Lyon) depuis 2013 et a enseigné auparavant à l'université d'Aix-Marseille, de Lausanne et de Cambridge. Ses travaux portent sur l'histoire de la pensée politique et économique des XVIII<sup>e</sup> et XIX<sup>e</sup> siècle, ainsi que sur la philosophie de l'économie. Il prépare actuellement un ouvrage sur la rationalité et l'identité (autour de Parfit) dans la collection Feuillet : économie politique moderne de l'ENS.

**Résumé** La fin du XIX<sup>e</sup> siècle est une période difficile pour les Français. La défaite de 1871, la chute du régime impérial, les déboires de la Troisième République et le krach de 1882 les ont amenés à douter de la capacité de leurs gouvernants. La société se fragilise et se divise, favorisant la montée du boulangisme et de l'antisémitisme. Charles Maurras et l'Action française en profitent également, après le suicide du général Boulanger en 1891 et la condamnation du capitaine Dreyfus en 1894. Comtien convaincu, Maurras prend régulièrement position contre Rousseau, dans lequel il voit le principal responsable de la décadence morale et politique de la France. S'ouvre alors une campagne d'une rare violence contre les valeurs morales et esthétiques du romantisme. Erigé en icône du romantisme par les partisans d'une « Nouvelle Renaissance Classique », Rousseau est accusé de tous les maux de la société moderne : l'individualisme exacerbé, la perversion des mœurs, le cosmopolitisme, le républicanisme, le sionisme, l'anarchisme et même l'homosexualité.

**Mots clés** Rousseau, Maurras, romantisme, Action française, théorie de la réception

**Abstract** *The end of XIX<sup>th</sup> century is a difficult time for the French people. The confidence they had in their politicians is shaken by Napoleon III's military defeat in 1871, the consecutive fall of his Empire, the political misfortunes of the new Republic or by the economic Krach of 1882. "Boulangism" and anti-semitism are the prime beneficiaries of the growing social and racial dissents in France. So are Charles Maurras and his Royalist party "L'Action Française" after the suicide of general Boulanger and the conviction of captain Dreyfus. Disciple of Auguste Comte, Maurras launches regular attacks on Rousseau whom he considers as responsible for the moral and political decadence of France. A violent campaign takes then place against the moral and aesthetic values of romanticism. Branded as an icon of romanticism by the defendants of a "New Classical Renaissance", Rousseau comes to embody all possible evils of modern society: excessive individualism, moral corruption, cosmopolitism, republicanism, Zionism, anarchism and even homosexuality.*

**Keywords** Rousseau, Maurras, Romanticism, Action Française, Reception theory

- 1 Cet article porte sur la réception de Rousseau pendant le premier tiers du XX<sup>e</sup> siècle en France. Une violente campagne contre le romantisme, incarné par

Rousseau, est alors menée par Maurras et ses disciples. En dépit de son caractère forcément arbitraire, j'ai retenu pour la périodisation de cette étude les bornes de 1898, date de la fondation de l'Action française, et de 1933, date de l'accès au pouvoir d'Hitler. Bien que la campagne menée par l'extrême droite contre Rousseau se poursuive dans les années trente et quarante, celle-ci prend une forme différente et elle répond à un autre contexte. Elle a, d'ailleurs, été déjà très bien étudiée<sup>1</sup>. Cette période constitue un moment critique dans l'histoire de la réception de Rousseau en France. L'objet de cette étude est d'expliquer les raisons pour lesquelles Rousseau a particulièrement été ciblé par ces attaques, et quelles étaient les motivations politiques réelles des sympathisants de l'Action française. D'un point de vue méthodologique, elle se propose de souligner l'importance de la prise en compte de la réception d'une œuvre dans l'histoire des idées. La première section porte sur le rejet culturel et politique du romantisme et le mouvement de Renaissance classique, dont Charles Maurras prend rapidement la tête au début du XX<sup>e</sup> siècle. La deuxième section se concentre sur les attaques menées contre Rousseau, qui en vient à représenter tous les dérèglements esthétiques et moraux du XIX<sup>e</sup> siècle, mais également – par extension – de ceux du début du XX<sup>e</sup> siècle. Les troisième et quatrième sections abordent deux points cardinaux de la campagne antirousseauiste : la personnalité du philosophe de Genève, et en particulier ses troubles physiques et psychiques, véritable allégorie de la « maladie » du romantisme et l'essence révolutionnaire et anarchique de sa pensée politique.

### La « Renaissance classique » : 1898-1914

- 2 Les lectures très critiques dont fait l'objet Rousseau à la fin du XIX<sup>e</sup> et au début du XX<sup>e</sup> siècle s'inscrivent dans une vague de contestation politique très violente à l'égard du romantisme. Pourtant, le romantisme est un mouvement littéraire qui a connu son apogée vers 1830. Il n'est pas politique : on y trouve aussi bien des écrivains socialistes tels que George Sand que des légitimistes tels que Chateaubriand. Surtout, le romantisme a périclité vers 1840. Le positivisme et le naturalisme qui lui ont succédé sont eux-mêmes sur le déclin vers 1890. Les attaques contre le romantisme de Rousseau ou de Chateaubriand sont une manière pour les conservateurs de s'en prendre aux nouvelles idoles intellectuelles du moment (Bergson et Freud notamment) ainsi qu'aux valeurs « corrompues » de la Belle Époque. « [A]ujourd'hui encore nous en sommes imprégnés jusqu'aux moelles. Il y a du romantisme, et à forte dose, dans la vie et dans la politique française », écrit Constantin Lecigne en 1909<sup>2</sup>. S'attaquer au romantisme est une façon de critiquer l'évolution des standards littéraires mais également celle des codes sociaux, en particulier la mode des dandys. Le romantisme, aux yeux de ses détracteurs, est un épiphénomène de tous les maux de la société. Selon Pierre Lasserre, il est tout à la fois « [R]uine psychique de l'individu, eudémonisme lâche, chimérisme sentimental, maladie de la solitude, corruption des passions, idolâtrie des passions, empire de la femme, empire des éléments féminins de l'esprit sur les éléments virils, asservissement au moi, déformation emphatique de la réalité, conception révolutionnaire et

<sup>1</sup> T. L'Aminot, « Jean-Jacques Rousseau face à la droite française (1940-1944) », *Studies on Voltaire and the Eighteenth Century*, n° 242, 1986, p. 473-489. Voir aussi P. Pellerin, *Les philosophes des Lumières dans la France des années noires : Voltaire, Montesquieu, Rousseau et Diderot 1940-1944*, Paris, L'Harmattan, 2009, ainsi que le chapitre consacré au totalitarisme dans C. Spector, *Au prisme de Rousseau : usages politiques contemporains*, Oxford, Voltaire Foundation, vol. VIII, 2011.

<sup>2</sup> C. Lecigne, *Le fléau romantique*, Paris, Lethellieux, p. 10.

à la Renaissance classique, il fait de Nietzsche l'un des meilleurs alliés de la cause de Maurras, et implore ce dernier d'être plus clément à son égard<sup>16</sup>. Peine perdue : Maurras ne partagera jamais l'admiration de Lasserre pour Nietzsche<sup>17</sup>. L'année suivante, toujours dans l'enquête du *Mercur de France*, Édouard Dujardin résume très justement le sentiment des Français envers Nietzsche : « Les temps changent. L'esprit allemand, par un admirable phénomène, vient de se nier lui-même en produisant le grand homme qui, allemand, représente la pure tradition française classique ; je veux parler de Nietzsche. »<sup>18</sup> Parmi les critiques du romantisme influencés par Nietzsche, citons également Ernest Seillière, qui développe une théorie originale associant volonté de puissance (qu'il appelle impérialisme), rejet du romantisme et tradition chrétienne<sup>19</sup>. Bien que proche des thèses maurrassiennes, il reproche à ce dernier son allégeance à Comte dont il considère le mysticisme comme un relent du romantisme.

- 9 C'est une crise politique, enfin, qui se caractérise par une vague d'antiparlementarisme boulangiste, puis royaliste, associée à une montée en puissance du nationalisme. La Ligue de la patrie française et l'Action française sont fondées en 1898 sur des positions explicitement nationalistes et antidreyfusardes. Les royalistes sont à la fin du XIX<sup>e</sup> siècle encore relativement peu nombreux, et sont essentiellement composés de catholiques conservateurs. La montée en puissance de Charles Maurras, dont les attaques contre le christianisme sont fréquentes, peut donc surprendre. Elle s'explique en grande partie par sa personnalité, très charismatique, mais également par sa philosophie morale et culturelle, très inspirée du positivisme comtien. Maurras rallie ainsi à sa cause tous ceux qui, bien que n'étant pas initialement royalistes, voient en lui le héraut d'un retour aux valeurs traditionnelles de la France. Maurras, pour qui le monothéisme favorise l'individualisme (et donc l'anarchisme), est d'ailleurs beaucoup plus sévère avec le protestantisme qu'il ne l'est avec le catholicisme, dont le monothéisme est – selon lui – limité<sup>20</sup>. Bien qu'agnostique, Maurras se retrouve donc paradoxalement à la tête d'un parti politique essentiellement composé de catholiques. Léon de Montesquiou est l'un des rares membres de l'Action française à le suivre sur cette voie. Cette position est critiquée par les chrétiens, notamment par Marc Sangnier qui, en 1904, remet en cause la compatibilité du positivisme et du catholicisme<sup>21</sup>. Bernanos, qui fut Camelot du roi (groupe de militants royalistes et antisémites) avant-guerre, et Jacques Maritain qui rompra avec Maurras après la condamnation de l'Action française par le pape Pie XI en 1926, comptent également parmi les dissidents catholiques de l'Action française. En dépit de ces critiques, la vision laïque du royalisme proposée par Maurras l'emporte. De ce point de vue, la séparation entre l'Église et l'État, promulguée en 1905, a probablement joué en sa faveur. Eugen Weber note d'ailleurs le nombre croissant de religieux qui, de dépit ou de colère, se tournent vers l'Action française en 1905-2006<sup>22</sup>.
- 10 L'autorité morale de Maurras parmi les membres et les sympathisants de l'Action française est, cependant, progressivement remise en cause. Tout d'abord, en plus

<sup>16</sup> *Id.*, « Charles Maurras et la Renaissance classique », art. cité, p. 612.

<sup>17</sup> C. Maurras, *Quand les Français ne s'aimaient pas. Chronique d'une renaissance, 1895-1905*, Paris, Nouvelle Librairie nationale, 1916.

<sup>18</sup> E. Dujardin, cité par C. Digeon, *La crise allemande de la pensée française*, ouvr. cité, p. 454.

<sup>19</sup> P. Dominique, *La pensée d'Ernest Seillière*, Paris, Nouveau Mercure, 1923 ; J. M. L. Boudeau, *Ernest Seillière. Historien du mysticisme romantique*, Paris, Émile-Paul Frères, 1925.

<sup>20</sup> M. Sutton, *Nationalism, Positivism, and Catholicism: The Politics of Charles Maurras and French Catholics, 1890-1914*, Cambridge, Cambridge University Press, 1982, p. 20.

<sup>21</sup> Charles Maurras lui répond dans *Le dilemme de Marc Sangnier : essai sur la démocratie religieuse* (Paris, Nouvelle librairie nationale) en 1906.

<sup>22</sup> E. Weber, *L'Action française*, trad. Michel Chrestien, Paris, Stock, 1964, p. 84.

dévergondée de la nature humaine, abus des moyens matériels de l'art pour masquer la paresse et la misère de l'invention»<sup>3</sup>.

- 3 Selon Léon Daudet, fils d'Alphonse et militant très actif de l'Action française, «[A]ucun régime, y compris celui de la Terreur, n'a été plus funeste à la France que celui qui va de l'automne de 1897 (début de la campagne dreyfusienne) à la première victoire de la Marne (septembre 1914)»<sup>4</sup>. Pour les antidreyfusards et les antiparlementaires, qui se réuniront bientôt au sein de la Ligue pour la patrie française ou dans l'Action française, la crise que traverse la France est triple : elle est morale, sociale et politique.
- 4 C'est une crise morale, tout d'abord, pour tous les conservateurs, catholiques ou traditionalistes, qui assistent au déclin des valeurs religieuses et au triomphe du dandysme. Elle fédère, par réaction, de nombreuses personnalités intellectuelles ou politiques telles que Georges Bernanos, Maurice Barrès, Paul Bourget, Charles Péguy et naturellement Charles Maurras. Tous voient dans la société moderne une décadence morale causée par la glorification de l'individualisme au détriment des valeurs de la religion, de la famille et de la nation. En 1902, Pierre Lasserre déjà désigne le romantisme comme responsable de la décadence morale et politique de la France :
- Dans tout ce qui est affaire d'estimation, de sentiment humain, les principes que la Grèce doit nous apprendre à mépriser le plus semblent avoir pour eux les vents et les courants. C'est, en politique, l'idolâtrie de la multitude, l'horreur des hiérarchies naturelles, l'incapacité d'obéir tant que de commander. C'est en morale l'abandon à toutes les impulsivités du cœur, aux pitiés désordonnées, un matérialisme sentimental, féminin, exprimé dans un langage d'idéalisme et de désintéressement hypocrites, la suppression de la souffrance individuelle donnée pour la fin suprême des sociétés ! En art, le romantisme vit toujours, non précisément à Paris – où depuis assez longtemps un chétif pullulement littéraire, sujet à peine à de brèves modes, ne suit plus aucune direction générale –, mais à Christiania, à Pétersbourg d'où il nous revient dépouillé de son éclat latin d'image et de pittoresque, accru de violence intérieure, assombri de frénésie prêchante.»<sup>5</sup>
- 5 Son œuvre majeure, et l'une des principales attaques contre le romantisme, *Le Romantisme français*, paraît en 1907. Le romantisme y est décrit comme un mouvement d'idées, une idéologie, davantage qu'un simple courant artistique. Son principe fondateur est celui de l'individualisme, responsable de la « ruine de l'individu.
- 6 Le romantisme, pour Lasserre, est indissociablement lié aux idéaux révolutionnaires de 1789 et 1793. C'est un courant moralement subversif et socialement destructeur. Psychologiquement, le romantisme est un dérèglement des sens, il est « primitivement maladie », assure Lasserre<sup>6</sup>. Il est, pour Ernest Seillière, qui publie *Le Mal romantique* en 1908, « dans son essence une insurrection du sentiment ou plutôt de l'instinct contre la raison »<sup>7</sup>. Il le qualifiera

<sup>3</sup> P. Lasserre, *Le Romantisme français. Essai sur la révolution dans les sentiments et dans les idées au XIX<sup>e</sup> siècle*, Paris, Société du Mercure de France, p. 311-312.

<sup>4</sup> L. Daudet, *Vers le roi*, cité par G. Kauffmann, « De Drumont à Maurras, une veine pamphlétaire », *Le maurrassisme et la culture. L'Action française. Culture, société, politique (III)*, O. Dard, M. Leymarie, N. McWilliam éd., Villeneuve d'Ascq, Presses universitaires du Septentrion, 2010, p. 28.

<sup>5</sup> P. Lasserre, « Charles Maurras et la Renaissance classique », *Mercure de France*, n° 150, juin 1902, p. 609-610.

<sup>6</sup> P. Lasserre, *Le Romantisme français*, ouvr. cité, p. 18.

<sup>7</sup> E. Seillière, *Le Mal romantique. Essai sur l'impérialisme irrationnel*, Paris, Plon-Nourrit, 1908, p. III.



également plus tard, en s'appuyant sur les travaux de psychologie expérimentale du docteur Pierre Janet, comme le symptôme d'une « dépression psychique morbide »<sup>8</sup>. L'analogie médicale est d'ailleurs reprise et développée dans le compte rendu que René Doumic fait de l'ouvrage de Lasserre et qu'il intitule *Pathologie du romantisme*<sup>9</sup>. Pour Doumic, le romantisme est une pathologie de l'ordre social qui « a pris à tâche de glorifier le paresseux, l'impuissant, le raté. Aventuriers de profession, escrocs, bandits, forçats, assassins, bouffons, truands, le caractère commun qu'il leur reconnaît, c'est la grandeur morale. »<sup>10</sup>

- 7 C'est une crise sociale, ensuite, dans laquelle l'affaire Dreyfus révèle au grand jour – au-delà de l'antisémitisme latent de la société française – la perte de l'autorité morale de l'État, et plus généralement celle de toute hiérarchie sociale. Elle se traduit par une radicalisation antidémocratique (Maurras, Sorel) ainsi que par la violence croissante des discours envers les minorités religieuses (notamment juives et protestantes), ainsi que contre les étrangers ou « métèques », en particuliers contre les Allemands accusés d'avoir corrompu la culture classique de la France avec le romantisme. Mais le véritable tournant vers le nationalisme antigermanique, selon Claude Digeon, a lieu avec la nouvelle politique expansionniste et colonialiste (la *Weltpolitik*) de l'empereur Guillaume II. Son discours du 31 mars 1905, qui inaugure ce que l'on a appelé la « crise de Tanger », annonce pour beaucoup la guerre à venir. De ce moment date, notamment, le retournement idéologique de Charles Péguy<sup>11</sup>. L'Allemagne devient alors l'ennemie absolue de la France, de son unité et de sa culture. Wagner et Kant, hier salués, sont conspués. Ceux qui admirent et adoptent leurs idées ou leurs goûts ont été corrompus et menacent désormais l'avenir de la culture et – à terme – de l'indépendance française. Le danger est important car, disent-ils, les valeurs allemandes sont déjà profondément enracinées en France : la musique, la littérature, la politique, la religion même sont dominées par l'influence allemande. Le mot d'ordre est celui de « renaissance », d'un retour aux valeurs traditionnelles et méditerranéennes de la France<sup>12</sup>.
- 8 Seul Nietzsche trouve faveur chez les nationalistes français. Bien qu'allemand, Nietzsche est, en effet, considéré comme un allié de la culture classique française par ses commentateurs. Son opposition à Wagner, et plus généralement au romantisme (à partir de 1878), son antiparlementarisme, son anticosmopolitisme et son antisémitisme le rendent sympathiques aux yeux des militants nationalistes français. Jacques Bainville, dont le premier article en 1897 est consacré à Nietzsche<sup>13</sup>, est l'un des premiers à faire son éloge et à voir en lui un allié de la cause française : « Je retiens les bons coups qu'il a portés à la détestable espèce des moralistes, à l'église humanitaire et à la gnose démocratique : ils ont fait réfléchir un certain nombre de Français, car c'est en France qu'ils trouvaient le mieux à s'appliquer. »<sup>14</sup> Pierre Lasserre, en 1902, qualifie Nietzsche de « Français par le goût »<sup>15</sup>. La même année, dans un article consacré

<sup>8</sup> *Id.*, Jean-Jacques Rousseau, Paris, Librairie Garnier, 1921, p. 152.

<sup>9</sup> R. Doumic, « Pathologie du romantisme », *Revue des Deux Mondes*, 15 avril 1907, p. 924-935.

<sup>10</sup> *Ibid.*, p. 930.

<sup>11</sup> C. Digeon, *La crise allemande de la pensée française (1870-1914)*, Paris, Presses universitaires de France, 1959, p. 505.

<sup>12</sup> Voir par exemple H. Clouard, *Les disciplines : nécessité littéraire et sociale d'une Renaissance classique*, Paris, M. Rivière, 1913.

<sup>13</sup> J. Bainville, « Un nouveau livre de Nietzsche », *L'Avenir artistique et littéraire*, 1<sup>er</sup> septembre 1897, p. 71-76.

<sup>14</sup> *Id.*, « Enquête sur l'influence allemande », *Mercure de France*, n° 155, novembre 1902, p. 299-300.

<sup>15</sup> P. Lasserre, *La Morale de Nietzsche*, Paris, Société du Mercure de France, p. 10.

d'être idéologiquement gênant, le positivisme de Maurras est philosophiquement dépassé et il a du mal à attirer la jeune génération tout entière vouée à Bergson. Son royalisme est également jugé inadapté à la société française moderne, et certains antiparlementaires préfèrent se tourner vers le syndicalisme révolutionnaire de Sorel. Une alliance entre ces deux mouvements est alors envisagée. Entre 1909 et 1914, on assiste ainsi à un surprenant rapprochement entre les royalistes de l'Action française et les syndicalistes révolutionnaires disciples de Georges Sorel. Un groupe de réflexion commun, le Cercle Proudhon, est créé sous l'impulsion du maurassien George Valois et du jeune sorellien Henri Lagrange. Ce rapprochement, qui aurait été rendu possible par la proximité physique des Camelots du Roi et des syndicalistes révolutionnaires à la prison de la Santé<sup>23</sup>, est fondé sur un rejet commun de la démocratie, d'une part, et du romantisme d'autre part. En lieu et place de la noblesse héréditaire de la monarchie, le système de Sorel, que Lasserre décrit comme un croisement du marxisme et du nietzschéisme<sup>24</sup>, instituerait une aristocratie syndicaliste.

- 11 Au niveau institutionnel, la montée des dissidences se traduit par la publication de revues « parallèles » à *L'Action française* alors dirigée par Vaugeois et Lasserre : création de la *Revue critique des idées et des livres* en 1908 par Jean Rivain pour défendre le néoclassicisme français, et de *L'Indépendance*, entre 1911 et 1913, par Jean Variot (pour Georges Sorel). Mais elles s'éloignent rapidement du « dogme » maurassien, en particulier en faisant une part trop importante aux romantiques et en préférant Bergson à Comte. En 1914, Rivain démissionne de l'Action française et le Cercle Proudhon, qui avait pour but la synthèse du monarchisme maurassien et du syndicalisme révolutionnaire de Sorel, est dissous. Gilbert Maire et Henri Clouard, rédacteurs de la *Revue critique*, prennent leurs distances avec Maurras pour défendre une approche plus ouverte du classicisme littéraire<sup>25</sup>. Enfin, en 1914, Lasserre se sépare également de Maurras et de l'Action française qu'il accuse de vouloir caricaturer sa pensée. Ses écrits postérieurs sont plus ouverts envers l'influence allemande et envers le romantisme. Dans sa *Jeunesse de Renan*, publié en 1925, il tente ainsi d'expliquer (plutôt que de condamner) l'intérêt intellectuel que Renan portait à l'Allemagne. Dans *Des Romantiques à nous* (1927), il atténue le rôle de Rousseau et reconsidère ses positions notamment vis-à-vis de la Révolution française et des auteurs Romantiques qui lui sont associés (Michelet, Hugo)<sup>26</sup>.

## Rousseau et la campagne contre le romantisme

- 12 Nous avons vu dans la section précédente que la critique du romantisme était constitutive du discours de l'Action française et de ses sympathisants. Cependant, la relation de la droite conservatrice et royaliste à l'égard du romantisme ne doit pas être schématisée. Derrière le discours officiel se cachent

<sup>23</sup> M. Leymarie, « Dissidents et critiques des années vingt », *L'Action française : culture, société, politique*, M. Leymarie et J. Prévotat éd., Villeneuve d'Ascq, Presses universitaires du Septentrion, 2008, p. 375.

<sup>24</sup> P. Lasserre, Georges Sorel, théoricien de l'impérialisme. Ses idées. Son action, *Cahiers de la Quinzaine*, vol. XVII, 1928.

<sup>25</sup> H. Clouard, « L'Aventure de l'art moderne. Sur une lettre de M. René Gillouin », *Revue critique des idées et des livres*, t. XXII, n° 130, 10 septembre 1913, p. 525-542. L'article est, en partie, repris dans H. Clouard, « Sur le programme des néo-classiques », *Revue critique des idées et des livres*, t. XXIV, n° 140, 10 février 1914, p. 269-282; G. Maire, « Précisions sur l'individualisme », *Revue critique des idées et des livres*, t. XXIII, n° 133, 25 octobre 1913, p. 129-141.

<sup>26</sup> W. M. Frohok, *Pierre Lasserre. The Evolution of his Critical Doctrines*, Ann Arbor, Edwards Brothers, 1937.

des goûts littéraires et des parcours individuels souvent très différents. Jacques Bainville est très marqué par l'influence de Carlyle<sup>27</sup>. Henri Clouard fait l'éloge de Lamartine, de Musset ou de Guérin<sup>28</sup>. Certains auteurs romantiques font d'ailleurs l'objet de disputes politiques ou littéraires. Chateaubriand, par exemple, honni par Maurras, fut ainsi un ardent légitimiste et catholique. Jacques Bainville en fait d'abord l'éloge, avant que sa proximité avec Rousseau (soulignée par Maurras) ne l'en éloigne<sup>29</sup>. La création d'un prix Stendhal par la *Revue critique*, qui lui consacre son numéro du 10 mars 1913, provoque l'ire de Maurras<sup>30</sup>. Un seul auteur fait l'unanimité contre lui, c'est Rousseau.

- 13 La véritable campagne de dénigrement contre Rousseau débute en 1907-1908. Elle est menée par deux grandes figures de l'extrême droite, Pierre Lasserre et Jules Lemaître. Le premier, rédacteur du journal *L'Action française*, est un proche de Charles Maurras. Le second est, avec Maurice Barrès, membre dirigeant de la Ligue de patrie française, fondée en 1898 en réaction à la création de la Ligue des droits de l'homme. Les auteurs romantiques sont accusés de faire de leur personne et de leurs sentiments l'objet unique de leur intérêt. Pour Paul Bourget, « [C]omme tous les dégénérés de son espèce, Rousseau est à la fois un égotiste et un émotif. Il ne voit que lui, ne connaît que lui. Son "moi" hypertrophié ne lui permet pas de se représenter dans sa réalité concrète une autre créature que lui »<sup>31</sup>. Pour Lemaître, Rousseau est d'un « individualisme outré »<sup>32</sup>. Le romantisme se définit encore, selon Lasserre, par « un parti d'individualisme absolu dans la pensée et dans le sentiment »<sup>33</sup>. Il consiste enfin, selon Émile Faguet dans son *Rousseau, penseur* (1910), « dans l'étalage du moi, le culte du moi et l'apothéose du moi »<sup>34</sup>. Rousseau, c'est la fin du classicisme, la fin du respect des règles esthétiques, morales ou politiques qui caractérisaient « l'ordre classique ». Tout cela a laissé place à la spontanéité de l'action, à la subjectivité des valeurs et à « ce culte stupide de l'égalité »<sup>35</sup> entre les hommes. « Il y avait », écrit John Charpentier, « une tradition qui pliait la matière aux lois de l'esprit avant l'auteur de *l'Émile* et du *Contrat*. Il l'a ruinée »<sup>36</sup>. C'est donc contre cette « ruine » de la culture française, réinventée et redynamisée par le naturalisme littéraire et le spiritualisme philosophique, que Maurras et ses sympathisants entrent en combat.
- 14 Bien qu'appartenant à deux mouvements politiques différents, Lasserre et Lemaître vont mener une campagne commune contre Rousseau et le romantisme. De calibres très différents, *Le romantisme français. Essai sur la révolution dans les sentiments et dans les idées au XIX<sup>e</sup> siècle* et *Jean-Jacques Rousseau*, leurs ouvrages respectifs, tous deux publiés en 1907, vont exercer une très forte influence jusqu'à la seconde guerre mondiale. Le livre de Lemaître connaît un très grand succès populaire. Il est immédiatement traduit en anglais

<sup>27</sup> C. Dickès, *Jacques Bainville : les lois de la politique étrangère*, Paris, Bernard Giovanangeli, 2008, p. 16-19.

<sup>28</sup> H. Clouard, « L'Aventure de l'art moderne », art. cité, p. 533.

<sup>29</sup> « Je vous suis notamment reconnaissant de m'avoir montré ce qu'a été Chateaubriand. Il m'était cher par son loyalisme et je m'étais accoutumé à distinguer en lui l'écrivain et le politique : en sorte que je ne lui reconnaissais pas littérairement sa parenté avec Rousseau et son anarchisme sentimental. » Cité par C. Dickès, *Jacques Bainville*, ouvr. cité, p. 39.

<sup>30</sup> G. DeLeonibus, « Discordances du classicisme maurassien », *Le maurassisme et la culture*, M. Leymarie et J. Prévotat éd., ouvr. cité, p. 168.

<sup>31</sup> P. Bourget, « Sur Jean-Jacques Rousseau », *La Revue critique*, vol. XVII, n° 101, 1912, p. 643.

<sup>32</sup> J. Lemaître, *Jean-Jacques Rousseau*, Paris, Calmann-Lévy, 1939, p. 338.

<sup>33</sup> P. Lasserre, *Le Romantisme français*, ouvr. cité, p. 17.

<sup>34</sup> E. Faguet, *Rousseau penseur*, Paris, Société française d'imprimerie et de librairie, 1910, p. 406.

<sup>35</sup> J. Lemaître, *Jean-Jacques Rousseau*, ouvr. cité, p. 124.

<sup>36</sup> J. Charpentier, *Jean-Jacques Rousseau ou le démocrate par dépit*, Paris, Perrin et C<sup>ie</sup>, 1931, p. 331.

et sera réédité en France presque sans discontinuité jusqu'en 1939. Albert Schinz note qu'il atteint sa trente-huitième édition en 1921<sup>37</sup>! Le travail académique de Lasserre est, quant à lui, considéré comme une référence. Il sera notamment salué par Carl Schmidt dans son *Politische Romantik* (1919). L'ouvrage de Lasserre s'ouvre par un chapitre consacré à Jean-Jacques Rousseau. Selon Lasserre, en effet, «Rousseau n'est pas à l'égard du romantisme un précurseur. Il est le romantisme intégral. [...] [R]ien dans le romantisme qui ne soit du Rousseau. Rien dans Rousseau qui ne soit Romantique»<sup>38</sup>. Cette posture, qui consiste à déconsidérer la totalité du mouvement romantique en l'associant à Rousseau (et inversement), est à l'époque également adoptée par deux autres importants critiques du romantisme, le néo-nietzschéen Ernest Seillière (qui publie *Le Mal romantique. Essai sur l'impérialisme irrationnel* en 1908 et *Jean-Jacques Rousseau* en 1921) et l'Américain Irving Babbitt (*Rousseau and Romanticism*, 1919). Dans son influente étude outre-Atlantique, Babbitt admet ainsi «qu'attaquer ou défendre Rousseau est le plus souvent simplement un moyen d'attaquer ou de défendre ce mouvement»<sup>39</sup>.

- 15 La campagne de l'Action française contre le romantisme franchit une nouvelle étape avec le bicentenaire de la naissance de Rousseau en 1912<sup>40</sup>. Les *Annales de la Société Jean-Jacques Rousseau* sortent, pour l'occasion, un numéro spécial qui comprend, notamment, un remarquable article de Gustave Lanson sur l'unité de la pensée de Rousseau ainsi que deux articles sur sa postérité, respectivement signés par Daniel Mornet (pour le XVIII<sup>e</sup> siècle) et Harald Höffding (pour le XIX<sup>e</sup> siècle). Parmi les célébrations nationales prévues figure notamment l'inauguration, le 30 juin, d'un monument à la gloire de Rousseau, œuvre du sculpteur Albert Bartholomé et situé au Panthéon. Elles provoquent de violentes réactions de la part de l'extrême droite, notamment des Camelots du roi. Les protestations et échauffourées du 30 juin sont le résultat d'une campagne conjointement menée par Maurice Barrès<sup>41</sup> et Charles Maurras contre les célébrations du bicentenaire. Raymond Trousson note que pas moins d'une trentaine d'articles sont ainsi publiés dans *L'Action française* entre juin et juillet 1912<sup>42</sup>. La conjonction du bicentenaire de Rousseau, de la dégradation de la situation étrangère et la montée de l'anarchisme en 1912 sont autant de prétextes pour les royalistes de l'Action française. Maurras reconnaît ainsi que «non seulement l'élite n'est plus «roussienne» aujourd'hui, mais son «antiroussisme» fait un beau chemin dans la masse. [...] tout le monde se rend compte que cette éclipse de l'intelligence française touche à sa fin. [...]»<sup>43</sup>
- 16 À l'approche de la «Grande guerre», la question de l'identité culturelle et religieuse de Rousseau prend une nouvelle dimension<sup>44</sup>. Tous ses critiques

**37** A. Schinz, *État présent des travaux sur Jean-Jacques Rousseau*, Paris, Les Belles Lettres, 1941, p. 64.

**38** P. Lasserre, *Le Romantisme français*, ouvr. cité, p. 14-15.

**39** I. Babbitt, *Rousseau and Romanticism*, Boston, Houghton, Mifflin, 1919, p. IX.

**40** Sur la commémoration de Rousseau en 1912, voir T. L'Aminot, *Images de Jean-Jacques Rousseau de 1912 à 1978*, Oxford, Voltaire Foundation, 1992, p. 23-60.

**41** Sur la relation parfois ambiguë qu'a entretenue Barrès avec Rousseau, voir T. L'Aminot, «Maurice Barrès et Jean-Jacques Rousseau. Histoire d'une relation», *Annales de l'Est*, 1983/2, p. 127-146.

**42** R. Trousson, *Rousseau et sa fortune littéraire*, Paris, A.-G. Nizet, 1977, p. 123.

**43** C. Maurras, «Le mal à la tête», *L'Action française*, 27 juin 1912, p. 1.

**44** En France, la fin du XIX<sup>e</sup> siècle est marquée, comme chacun le sait, par deux grands événements, la défaite de 1870 et l'affaire Dreyfus. Ces deux événements sont généralement cités pour expliquer le nationalisme très affirmé d'une grande partie des Français, y compris des intellectuels et des artistes, avant la première guerre mondiale. Claude Digeon montre cependant que la situation est plus complexe qu'il n'y paraît, et que la réaction nationaliste

s'accordent à voir en lui un « étranger » ou un émissaire de l'étranger. Le romantisme est, rappelle-t-on à l'envi, un mouvement littéraire de tradition allemande. Rousseau ne peut donc appartenir à la culture française. Il est, tout d'abord, d'origine genevoise. Il est, surtout, de culture protestante. Cela suffit à faire de lui un « métèque »<sup>45</sup>. Le protestantisme est, ainsi qu'on l'a mentionné plus haut, condamnable à deux titres pour Maurras. Il est, tout d'abord, l'expression d'une culture étrangère et même ennemie, celle de l'Allemagne. Il est, ensuite, la forme la plus achevée du monothéisme chrétien, qui place l'individu et son bien-être au centre de la relation à Dieu. À cet égard, le protestantisme est aussi dangereux pour la société que peut l'être le romantisme. Ainsi, lorsqu'il s'agit d'évoquer le très peu romantique *Contrat social*, c'est l'influence pernicieuse du protestantisme qui est mise en cause :

Rousseau est-il fou, ainsi qu'on l'a affirmé, lorsqu'il a composé ce chapitre du *Contrat social* où il immole l'indépendance du citoyen à l'omnipotence de l'État, où il sacrifie la souveraineté de la justice à la souveraineté de la majorité, où il se montre plus exclusif, plus intolérant que les théoriciens du despotisme d'État dans l'antiquité ? Non, il n'était pas fou, pas entièrement fou du moins, il était de tradition genevoise et calviniste.<sup>46</sup>

En définitive, les Droits de l'homme, proclamés en France, grâce à l'influence de Rousseau, étaient surtout les Droits de l'Allemand. [...] Que les Allemands le fêtent donc tout à leur aise. Mais nous, nous n'avons pas de dette envers lui.<sup>47</sup>

Il était bien plus qu'un Suisse, c'était la Suisse, ou, pour mieux dire, c'était toute la Germanie, toute l'Europe anglo-saxonne, ramenée à la sauvagerie intellectuelle et morale par la Réforme, c'était la NOUVELLE JUDÉE instituée par Luther et Calvin, que répétait la prédication de Rousseau.<sup>48</sup>

- 17 L'appartenance de Rousseau à la culture germanique, établie à travers le romantisme, va également jouer un rôle important dans la critique du nazisme par l'extrême droite française<sup>49</sup>. Bien que partageant les thèses antiparlementaires et antisémites du national-socialisme, Maurras et les membres de l'Action française rejettent la menace politique et culturelle que représente le nazisme. Ils considèrent surtout que le nazisme est une doctrine messianique et narcissique héritée de l'idéalisme allemand. Ils sont donc parmi les premiers à établir un lien entre rousseauisme et national-socialisme. Dès 1933, Maurras conspu les sympathisants « rousseauistes » du national-socialisme : « Comme Rousseau s'était reconnu dans un certain miroir allemand, comme, à leur tour, les Allemands se reconnurent dans Rousseau, voici que nos « roussiens » français se reconnaissent naturellement dans Hitler. »<sup>50</sup> En 1934, Gonzague de Reynold, catholique nationaliste proche de l'Action française, qualifie le *Contrat social* « d'ouvrage national-socialiste »<sup>51</sup>.

et surtout antigermanique apparaît beaucoup plus tardivement. La comparaison des enquêtes sur les relations intellectuelles franco-allemandes, publiées par le *Mercure de France* respectivement en 1895 et 1902-1903 et étudiées par Claude Digeon, est édifiante. C. Digeon, *ouvr. cité*, p. 463 et suiv.

**45** Voir, par exemple, « Le métèque Rousseau » [proclamation de « l'Action française », citations de Proudhon, Lemaître, Maurras], *Action française*, 28 juin 1912 ; L. Daudet, « La chienlit de Rousseau, Le Fou et les Singes », *L'Action française*, 30 juin 1912 ; M. Pujol, « À bas le métèque », *L'Action française*, 1<sup>er</sup> juillet 1912.

**46** A. Dide, *Jean-Jacques Rousseau, le protestantisme et la Révolution française*, Paris, Flammarion, p. 169.

**47** J. Bainville, « Rousseau et nous », *L'Action française*, 18 juin 1912, p. 1.

**48** C. Maurras, « Un ennemi de la France », *L'Action française*, 13 juin 1912, p. 1.

**49** T. L'Aminot, « Jean-Jacques Rousseau face à la droite française (1940-1944) », *Studies on Voltaire and the Eighteenth Century*, n° 242, 1986, p. 473-489.

**50** C. Maurras, « La politique », *L'Action française*, 1<sup>er</sup> octobre 1933, p. 1

**51** P. Pellerin, *Les philosophes des Lumières dans la France des années noires : Voltaire, Montesquieu, Rousseau et Diderot, 1940-1944*, Paris, L'Harmattan, 2009, p. 129.

## Rousseau, le « malade »

- 18 Le second grief fait à Rousseau est d'ordre moral. En rejetant les valeurs catholiques et traditionnelles de la France d'Ancien Régime, et en laissant les individus s'abandonner à leurs instincts naturels, Rousseau aurait encouragé le vice. « Dans l'ordre social », écrit Lecigne en 1909, « le romantisme c'est la réhabilitation des gueux, la chimère substituée au bon sens, la littérature de l'anarchie et du sentimentalisme révolutionnaire, – et pour trouver la formule première de cet évangile nouveau, il faut remonter jusqu'au *Contrat social* de Rousseau. »<sup>52</sup>. Selon une thèse toute nietzschéenne, Rousseau renverse les normes morales de son époque pour mieux les conformer à son propre caractère. La philosophie romantique trouverait ses origines dans la volonté de quelques individus, dont Rousseau, de normaliser leurs déviances personnelles, de renverser les codes de la moralité à leur profit. « Cet enfant de Rousseau », écrit ainsi Maurras, « met en bas ce qui était en haut, et inversement. »<sup>53</sup> François Mauriac est encore plus explicite :
- Jusqu'au jour du citoyen de Genève, les assassins se connaissaient en criminels, les libertins ne se donnaient pas en exemple, les sodomites n'enseignaient pas la morale [...]. Jean-Jacques peut reprendre le mot du *Médecin malgré lui* : « Nous avons changé tout cela. » Le cœur est à droite, le foie à gauche.<sup>54</sup>
- 19 Selon ses détracteurs, le romantisme correspond à ce que Nietzsche appelait une « morale de l'utilité ». Le romantisme est donc présenté comme une philosophie de malades créée par un malade, Rousseau.
- 20 Dans les années dix et vingt, la « maladie » de Rousseau est davantage qu'une question d'érudition pour les commentateurs de Rousseau. Elle constitue un élément essentiel du discours antiromantique : la pathologie du romantisme reflète celle de son fondateur. Rousseau est donc presque systématiquement présenté comme un malade. La « morbidité », c'est-à-dire le caractère maladif ou malsain, de Rousseau et de sa pensée (l'un et l'autre sont indissociables pour les commentateurs) est un élément récurrent de la littérature alors consacrée à Rousseau. « À la base de l'individualité morale de Rousseau », déclare ainsi Lasserre, « et l'affectant dans toutes ses parties, il y a un élément morbide »<sup>55</sup>. De même, pour Jacques Maritain, « [L]a pensée actuelle, en ce qu'elle a de morbide, reste encore sous sa dépendance. »<sup>56</sup>
- 21 Malade, Rousseau l'était certainement. Ses pathologies physiques, notamment ses troubles urinaires, sont bien connues. Rousseau en fait état dans ses *Confessions*, ou plutôt selon Lecigne, « [I]l a cru bon de nous confier toutes ses misères, de s'étaler devant nous comme un débris d'hôpital »<sup>57</sup>. Les pathologies de Rousseau (troubles urinaires, insomnie, troubles auditifs...) ont, par ailleurs, été souvent étudiées<sup>58</sup>. La maladie physique est généralement excusable. Chacun peut en être, en effet, victime. Dans le cas de Rousseau, cependant, les troubles

<sup>52</sup> C. Lecigne, *Le fléau du romantisme*, ouvr. cité, p. 67.

<sup>53</sup> C. Maurras, *Romantisme et Révolution. L'œuvre de Charles Maurras*, vol.3, Paris, Nouvelle Librairie nationale, 1922, p. 12.

<sup>54</sup> F. Mauriac, *Trois grands hommes devant Dieu*, Paris, Le Capitole, 1930, p. 67.

<sup>55</sup> P. Lasserre, *Le Romantisme français*, ouvr. cité, p. 20.

<sup>56</sup> J. Maritain, *Trois réformateurs : Luther, Descartes, Rousseau*, Paris, Librairie Plon, 1925, p. 164.

<sup>57</sup> C. Lecigne, *Le fléau romantique*, ouvr. cité, p. 49.

<sup>58</sup> Voir notamment L.-A. Mercier, *Explication de la maladie de Jean-Jacques Rousseau et de l'influence qu'elle a eue sur son caractère et sur ses écrits*, Paris, Le Normant, 1859 ; F. Girardet, *La mort de Jean-Jacques Rousseau : étude médicale*, A. Maloine, 1909 ; S. Elozu, *La maladie de Jean-Jacques Rousseau*, Paris, Librairie Fischbacher, 1929.



urinaires dont il souffrait sont rapportés à ses mœurs « dissolues ». Ainsi, pour Charpentier, « [S]’il l’entretient [sa maladie] par sa façon de vivre, ses pratiques onanistes l’aggravent »<sup>59</sup>. D’autres voient dans ce dysfonctionnement physique l’origine de certains de ses dysfonctionnements psychiques. Jules Lemaître émet ainsi l’hypothèse (qui sera ensuite reprise par Mauriac) que, devenu impuissant, Rousseau aurait tenté de le cacher en « inventant » l’histoire des cinq enfants abandonnés à l’hospice public<sup>60</sup>.

- 22 Mais ce qui intéresse le plus les détracteurs de Rousseau à l’époque, ce sont ses pathologies psychiques. La psychiatrie est alors, faut-il le rappeler, tout juste implantée en France, et le « cas Rousseau » est souvent donné en exemple. Les adversaires du romantisme peuvent donc aisément appuyer leur thèse sur des rapports et articles médicaux consacrés à Rousseau. Les travaux du docteur Pierre Janet, professeur de psychologie expérimentale au Collège de France, du docteur Régis<sup>61</sup>, professeur à Bordeaux, du docteur Möbius<sup>62</sup> (spécialiste de l’hystérie et père de la psychothérapie), du docteur Poncet<sup>63</sup>, professeur à Lyon, du docteur Cabanès<sup>64</sup> ainsi que du docteur Magnan sont ainsi sollicités par Seillière, Lasserre, Bourget et Proal<sup>65</sup> pour démontrer la « dépression psychique morbide »<sup>66</sup> ou « la neurasthénie spasmodique obsédante »<sup>67</sup> de Rousseau. Dans son article de 1912, Paul Bourget se fait ainsi fort d’accuser le gouvernement de non seulement glorifier un « malade » mais de promouvoir le système moral et social né de sa folie :

Les médecins peuvent différer d’opinion sur la lésion qui fut l’origine du désordre. Ils sont d’accord sur ce point qui seul importe : l’auteur des *Confessions* fut un malade mental, et il le fut toute sa vie. C’est un malade mental que les pouvoirs publics se proposent d’honorer comme un des prophètes de la Révolution, le plus efficace peut-être. Ils n’aperçoivent pas qu’ils jugent ainsi la nature de ce mouvement. Ce véritable accès de psychose collective devait avoir pour initiateur un psychopathe caractérisé.<sup>68</sup>

- 23 Enfin, parmi ce que l’on considère à l’époque comme une « affection » psychique, mentionnons la passivité et la féminité attribuée à Rousseau. Aujourd’hui considéré comme un auteur misogyne, Rousseau est avant-guerre considéré comme un « homme femme » ou un « femmelin », selon le mot de Proudhon, dont le détestable ouvrage est republié en 1912 avec une introduction de Henri Lagrange<sup>69</sup> :

D’inavouables idées baissent sa tête vers la terre. Son âme dérobe aux témoins une vie d’émotions fébriles. [...] Les précoces étrangetés de sensation et de désir qu’il nous avoue avoir été la hantise de son enfance ont leur source au fond de lui-même,

<sup>59</sup> J. Charpentier, *Jean-Jacques Rousseau ou le démocrate par dépit*, Paris, Perrin & Cie, 1931, p. 98.

<sup>60</sup> J. Lemaître, *Jean-Jacques Rousseau*, ouvr. cité, p. 59. Mauriac envisage également cette hypothèse : « l’inconfort de Rousseau le privait *depuis longtemps* de tout rapport sexuel. Or il n’est pas de crime dont un impuissant ne se chargerait pour dissimuler son malheur. [...] Ainsi nous paraît vraisemblable cette invention de Rousseau. » F. Mauriac, *Trois hommes devant Dieu*, ouvr. cité, p. 84-85 (Mauriac souligne).

<sup>61</sup> E. Régis, *La neurasthénie de Jean-Jacques Rousseau*, Bordeaux, Imprimerie Gounouilh, 1900 ; E. Régis, *Le testament de Jean-Jacques Rousseau de 1763*, Paris, Lecène-Oudin, 1907.

<sup>62</sup> P. J. Möbius, *Jean-Jacques Rousseau’s Krankheitsgeschichte*, Leipzig, Vogel, 1889.

<sup>63</sup> A. Poncet, *La maladie de Jean-Jacques Rousseau*, Paris, Masson, 1907.

<sup>64</sup> A. Cabanès, *Les grands névropathes*, vol. II, Paris, Albin Michel, 1931.

<sup>65</sup> L. Proal, *La psychologie de Jean-Jacques Rousseau*, Paris, Librairie Félix Alcan, 1923.

<sup>66</sup> E. Seillière, *Jean-Jacques Rousseau*, ouvr. cité, p. 151.

<sup>67</sup> P. Bourget, « Sur Jean-Jacques Rousseau », *Revue critique des idées et des livres*, vol. XVII, n° 101, 1912, p. 646. L’expression est empruntée au docteur Régis.

<sup>68</sup> *Ibid.*, p. 643.

<sup>69</sup> P. J. Proudhon, *Les femmelins*, Paris, Nouvelle Librairie nationale, 1912.

tels les gaz émanés d'un cadavre. Ce ne sont pas crises d'un âge, mais égarements innés de la sensibilité que le premier objet met à nu. Joignant à des enfantillages de collégien une science corrompue de vieillard, cette sensibilité ne sera jamais assainie par une vigoureuse onde virile.<sup>70</sup>

C'est la prédominance de la sensibilité qui donne aux jeunes gens tant de charmes, ainsi que l'heureuse confiance qu'ils ont en eux-mêmes; ils font honneur à leurs âmes des émotions qu'ils tiennent de leur âge, et les hommes sensibles ressemblent également aux femmes par beaucoup de traits. Or Rousseau avait à un degré éminent ce genre de sensibilité, tout ensemble *ardente et faible*; elle l'a servi dans ses ouvrages et égaré dans son existence, ainsi qu'elle égare après lui ses lecteurs.<sup>71</sup>

Il va « comme la tête lui chante », cueillant et effeuillant les heures au cours d'un nonchalant vagabondage, où sa jeunesse se prolonge jusqu'à s'efféminer... Cette jeunesse, il est vrai, n'est point virile. Les voluptés dont elle l'a comblé l'alanguissent : il ne les a pas voulues. Il les subit. C'est une manière d'androgynie que *Les Confessions* nous révèlent, et je m'explique, en dehors même d'une certaine répugnance, le mépris qu'elles nous inspirent en nous séduisant.<sup>72</sup>

Au lieu de dire que le romantisme a fait dégénérer les âmes ou les esprits français, ne serait-il pas meilleur de se rendre compte qu'il les effémina ?<sup>73</sup>

- 24 La confusion entretenue entre sensibilité et féminité, voire homosexualité (celle-ci fut « établie » en 1927 par René Laforgue<sup>74</sup>, l'un des pionniers de la psychanalyse en France), joue un rôle essentiel dans la campagne menée contre la décadence romantique. Ses adversaires lui opposent la grandeur et la virilité de la tradition classique. Le choix de l'auteur de *l'Émile* peut certes étonner : il est, en fait, dicté par l'identité posée par Lasserre en 1907 entre Rousseau et le mouvement romantique. Car derrière le mouvement romantique se profile tous les écrivains qui ont été parfois désignés comme néoromantiques, dont la « sensibilité » choquait les catholiques et les conservateurs contemporains :

L'aïeul Jean-Jacques est plus jeune que son fils Chateaubriand et que ses petit-fils, les romantiques. Eux dorment, embaumés dans leur gloire. Lui, il est l'un de nous – je finis sur ce mot par quoi j'ai commencé – : il s'appelle Romain Rolland, Marcel Proust, André Gide.<sup>75</sup>

## Rousseau, le révolutionnaire et l'anarchiste

- 25 Le troisième grief fait à Rousseau est de nature politique. La philosophie romantique est, selon ses détracteurs, anarchiste et révolutionnaire. La thèse de la responsabilité de Rousseau dans la Révolution Française n'est pas nouvelle. Burke, Maistre et Bonald l'ont très vite défendue. Les attaques politiques contre Rousseau au début du XX<sup>e</sup> siècle diffèrent, cependant, de celles du siècle précédent. Elles sont, tout d'abord, marquées par le spectre de la Commune de 1870 et de la menace communiste. Il n'est donc pas anodin à l'époque de voir en Rousseau « un socialiste, et même, au point de vue du génie, comme le plus grand des socialistes »<sup>76</sup>, ou encore d'affirmer que « [L]e communisme est enveloppé dans Rousseau »<sup>77</sup>. Intellectuellement, ensuite, elles sont très inspirées

<sup>70</sup> P. Lasserre, *Le Romantisme français*, ouvr. cité, p. 23.

<sup>71</sup> E. Seillière, *Jean-Jacques Rousseau*, ouvr. cité, p. 162 (Seillière souligne).

<sup>72</sup> J. Charpentier, *Jean-Jacques Rousseau*, ouvr. cité, p. 298.

<sup>73</sup> C. Maurras, « Le Romantisme féminin », art. cité, p. 189.

<sup>74</sup> R. Laforgue, « Étude sur Jean-Jacques Rousseau », *Revue française de psychanalyse*, vol. 1, n° 2, 1927; étude reprise et développée dans R. Laforgue, *Psychopathologie de l'échec*, Paris, Payot, 1939 (chap. 9).

<sup>75</sup> F. Mauriac, *Trois grands hommes devant Dieu*, ouvr. cité, p. 93.

<sup>76</sup> E. Faguet, *Rousseau penseur*, ouvr. cité, p. 282.

<sup>77</sup> J. Lemaître, *Jean-Jacques Rousseau*, ouvr. cité, p. 255.



par la philosophie de Nietzsche. Ainsi c'est la perspective de l'anarchie et du nihilisme qui est agitée face aux principes démocratiques de Rousseau. Ainsi pour Lasserre, « [L]es théories de Jean-Jacques sont la glorification de ses mœurs. [...] Elles aboutissent par cinq ou six voies différentes au nihilisme social qui est le vœu profond de son cœur. Elles proposent avec éloquence soit l'anarchie, soit la Révolution éternelle »<sup>78</sup>. Pendant le premier tiers du XX<sup>e</sup> siècle, les attaques politiques contre Rousseau se concentrent, donc, sur deux points spécifiques, la responsabilité morale qu'il porte à l'égard de la Révolution française et son influence sur l'anarchisme contemporain.

- 26 La question de la responsabilité de Rousseau dans la Révolution française est ancienne. Burke et Maistre déjà l'avaient soulevée, la Révolution à peine terminée. Mais le contexte politique français ayant changé après 1830, le paradigme d'une Révolution jacobine et rationaliste, porté par Tocqueville et Taine, s'était progressivement imposé. Faire de Rousseau un précurseur de la Révolution française à la fin du XIX<sup>e</sup> siècle constitue donc un renversement de perspective notable. Celui-ci s'explique naturellement par la volonté de Maurras et de ses partisans d'exonérer la tradition classique française de toute responsabilité dans ce qui est encore largement considéré comme une tache dans l'histoire de France. Cela s'explique également par la réévaluation de la place de Rousseau dans le mouvement romantique. Alors que le romantisme avait jusqu'alors été considéré comme un mouvement du XIX<sup>e</sup> siècle, qui avait Rousseau comme précurseur, la mode est désormais de voir dans le romantisme un mouvement qui débute au XVIII<sup>e</sup> siècle et dont Rousseau est l'une des figures majeures<sup>79</sup>. La perspective peut donc maintenant être renversée. Notons que la relation de Rousseau à la Révolution constitue l'un des points d'achoppement entre les doctrines de Maurras et de Sorel. Pour ce dernier, qui fait un compte rendu de l'ouvrage de Lemaître<sup>80</sup>, la Révolution n'est pas l'œuvre de Rousseau<sup>81</sup>. Toute sa théorie politique, explique-t-il, consiste au contraire à rétablir un système de type féodal.
- 27 Maurras, on le sait, a régulièrement associé Rousseau aux « désordres » de la Révolution française. Il déclare notamment que « Rousseau a été [...] la cause formelle de la Révolution ; il en a été l'âme et le génie, excitant les petits, stupéfiant et endormant les grands, donnant à l'attaque révolutionnaire des forces, à la défense traditionnelle de la faiblesse »<sup>82</sup>. Il trouve en Lemaître un important soutien. Selon ce dernier, en effet, Rousseau a une responsabilité directe dans les troubles révolutionnaires :

Le jargon révolutionnaire, c'est la langue de Rousseau mal parlée. [...] Rousseau est simplement, pour les nigauds et les coquins de ce temps-là, le sauveur, le rédempteur de l'humanité. Sans lui, sans quelques phrases de cet étranger dans son *Discours sur l'inégalité*, surtout sans son *Contrat social* (auquel il tenait si peu), il est possible qu'on n'eût pas songé, en 1792, à faire la république.<sup>83</sup>

<sup>78</sup> P. Lasserre, *Le Romantisme français*, ouvr. cité, p. 57.

<sup>79</sup> Voir notamment D. Mornet, *Le Romantisme en France au XVIII<sup>e</sup> siècle*, Paris, Librairie Hachette, 1912.

<sup>80</sup> G. Sorel, « Jean-Jacques Rousseau », *Le Mouvement socialiste*, juin 1907, p. 507-532.

<sup>81</sup> J. Julliard, « Rousseau, Sorel et la Révolution française », *Cahiers Georges Sorel*, 1985, vol. III, n° 3, p. 5-15.

<sup>82</sup> C. Maurras, *L'Action française*, 19 juin 1912, cité par R. Trousson, « Un ennemi de Rousseau : Charles Maurras », dans R. Trousson, *Défenseurs et adversaires de Jean-Jacques Rousseau. D'Isabelle de Charrière à Charles Maurras*, Paris, Honoré Champion, 1995, p. 318. Voir également, R. Trousson, *Rousseau et sa fortune littéraire*, Paris, A.-G. Nizet ; M. Weyembergh, *Charles Maurras et la Révolution française*, Paris, Vrin, 1992.

<sup>83</sup> J. Lemaître, *Jean-Jacques Rousseau*, ouvr. cité, p. 346-347.

- 28 La thèse de Lemaître et de Maurras, sans cesse martelée mais jamais véritablement discutée, est contestée en 1909 par Edmé Champion ainsi que par Henri Rodet. Rodet admet que la Révolution française de 1789 (tout comme celle, d'ailleurs, de 1848) puisse avoir été inspirée par la doctrine de la souveraineté populaire absolue de Rousseau. Mais, dans le même temps, il réfute toute intention malveillante de Rousseau dont le « cœur foncièrement bon et généreux, meurtri par ses plaies intimes et par la destinée, s'émouvait profondément au spectacle qu'il évoquait de la société contemporaine et de la condition misérable du plus grand nombre des hommes »<sup>84</sup>. L'ouvrage est immédiatement critiqué par Émile Faguet<sup>85</sup>. La monographie de Champion, *Rousseau et la Révolution française*, quant à elle, démonte pas à pas l'idée d'un Rousseau révolutionnaire, produit – selon l'auteur – d'une combinaison d'ignorance et de mécompréhension de la part de ses commentateurs. Il conclut qu'il « ne faut pas plus lui attribuer la Constitution civile du clergé que le culte de l'Être suprême »<sup>86</sup>. L'ouvrage de Champion suscite un fort intérêt. Deux études lui répondent en 1912. Le livre d'Auguste Dide, tout d'abord, intitulé *Rousseau, le protestantisme et la Révolution*, étudie en détail l'influence intellectuelle et personnelle que joua Rousseau sur Robespierre. S'il admet que la Révolution « laïque », c'est-à-dire celle de 1789, doit peu à Rousseau, en revanche il confirme que celle de 1792, dictée par son idéal de religion civile, est tout entière de son fait<sup>87</sup>. Celui de Meynier, *Rousseau révolutionnaire*, se présente (en dépit de son titre) comme une étude modérée et impartiale. Meynier y conclut que « l'influence de Rousseau fut, d'un bout à l'autre de la Révolution, réelle, mais bien inégale, intermittente et partielle, ne s'appliquant pas à la totalité des événements, restreinte au regard de leur complexité infinie, admissible certes, mais avec toute sorte d'atténuations »<sup>88</sup>. Signalons enfin la réédition opportune en 1926 de l'essai de Lamartine, *Jean-Jacques Rousseau : son faux Contrat social*, dans lequel le poète avançait que « [L]e Contrat social est le livre fondamental de la révolution française »<sup>89</sup>.
- 29 En 1912, cependant, ce n'est pas la Révolution mais bien l'anarchie qui concentre l'essentiel des attaques contre Rousseau<sup>90</sup>. Les circonstances sont alors propices. Les auteurs exploitent l'émotion soulevée par les braquages et les meurtres perpétrés depuis décembre 1911 par Jules Bonnot, Octave Garnier et Raymond Callemin, principaux membres du groupe anarchiste dit de la « bande à Bonnot ». Lors d'un discours prononcé à la Chambre des députés le 11 juin 1912, Maurice Barrès s'en prend ainsi au gouvernement radical-socialiste d'Armand Fallières qu'il accuse de soutenir les milieux anarchistes<sup>91</sup>. Barrès établit notamment un rapprochement entre la célébration nationale de Rousseau et la glorification des anarchistes Jules Bonnot et Octave Garnier, tués quelques semaines auparavant,

**84** H. Rodet, *Les idées politiques de Jean-Jacques Rousseau*, Paris, Arthur Rousseau, 1909, p. 428.

**85** E. Faguet, « La politique de Jean-Jacques Rousseau », *La Revue des Deux Mondes*, 15 septembre 1909, p. 396-409.

**86** E. Champion, *Jean-Jacques Rousseau et la Révolution française*, Paris, Armand Colin, 1909, p. 233.

**87** A. Dide, *Rousseau, le protestantisme et la Révolution*, ouvr. cité, p. 288.

**88** A. Meynier, *Jean-Jacques Rousseau révolutionnaire*, Paris, Schleicher Frères, 1912, p. 22.

**89** A. de Lamartine, *Jean-Jacques Rousseau : son faux Contrat social*, Paris, André Delpeuch, 1926, p. 70.

**90** Sur les lectures « anarchiques » de Rousseau en 1912, voir T. L'Aminot, *Images de Jean-Jacques Rousseau de 1912 à 1978*, Oxford, Voltaire Foundation, 1992.

**91** T. L'Aminot, « Les saligauds de la célébration. La bande à Bonnot, Rousseau et le "culte de la charogne" », *Études Jean-Jacques Rousseau*, n° 18, 2010-2011 : Rugosité de Rousseau, p. 153-178. Voir aussi T. L'Aminot, « Rousseau et la pensée anarchiste », *Frihetens Århundre*, n° 6, 2004, p. 176-207.

et qu'il présente comme « ses disciples ». La responsabilité morale de Rousseau dans les agissements sanglants de la « bande à Bonnot » est à nouveau invoquée, deux semaines plus tard, dans un article de *L'Action française* intitulé « Rousseau métèque. Bonnot, Garnier et Flachon au Panthéon »<sup>92</sup>, ainsi que dans la revue *L'Univers*<sup>93</sup>. Léon de Montesquiou consacre un article entier (et en première page) à « l'anarchiste » Rousseau :

Oui, Rousseau, recherchant une société peut arriver à s'affranchir de ses traditions, appelle de ses vœux pour cela un de ces moments de révolution où, dit-il, « l'horreur du passé tient lieu d'oubli ». Vœu impie que la Révolution réalisera, on sait comment, et qui conduit naturellement et logiquement à la haine de la Patrie. [...] Mais, direz-vous, tout cela, c'est l'anarchie. Oui, tout cela, c'est l'anarchie, car tout cela nous souffle la révolte contre toute autorité, qu'elle soit politique, sociale ou spirituelle, contre la société, la famille, la Patrie, la tradition. Oui, c'est l'anarchie autant qu'il se peut.<sup>94</sup>

30 En subvertissant le modèle social français, Rousseau aurait donc détruit les fondements traditionnels que sont la famille et la nation au profit des seuls individus et, de ce fait, posé les fondements de l'anarchie moderne.

31 On peut s'étonner à bon droit du fait que Rousseau, auteur du *Contrat social*, ait pu incarner le rejet de l'État et de la loi. Ce point est très révélateur du poids disproportionné alors accordé à son œuvre littéraire au regard de son œuvre politique. Mauriac le note d'ailleurs encore en 1942 :

Le malheur de Rousseau est que si les *Confessions* et les *Rêveries* nous demeurent familières, il n'en va pas de même pour le reste de son œuvre. Si nous relisons le *Contrat social* ou *l'Émile*, nous verrions qu'il s'est fait à lui-même plusieurs objections qu'on lui a opposées dans la suite; il n'eût certes pas souscrit à telles suppositions qu'on lui attribue, ni aux caricatures que ses contradicteurs donnent de sa pensée. Mais comme nous ne relisons que l'ouvrage immortel où il s'est mis à nu, où il se livre tel qu'il était, notre mauvaise humeur donne la mesure de la vérité du portrait.<sup>95</sup>

32 Robert Derathé fait une remarque similaire en 1948<sup>96</sup>.

33 Notons que ses opposants ont d'ailleurs tout intérêt à minimiser ou à discréditer le *Contrat social*. L'ouvrage est, tout d'abord, incompatible avec une lecture romantique de Rousseau. Il déplaît donc aux partisans de la Renaissance classique et du positivisme. Dans le *Contrat social*, ensuite, Rousseau ne s'oppose absolument pas à l'idée d'un gouvernement monarchique (en particulier en France où il est même, selon lui, approprié). Seule la souveraineté démocratique constitue un impératif politique. Royalistes et antiromantiques préfèrent donc l'ignorer ou en montrer l'incohérence. Émile Faguet le considère ainsi comme « un

<sup>92</sup> « Cette année 1912 a vu tomber deux centenaires : le centenaire de *Jeanne d'Arc* et celui de *Jean-Jacques Rousseau*. Le premier était la fête d'une *Française*, celle qui sauva la Patrie. [...] Le second était le centenaire d'un *Étranger*, celui qui fit le plus de mal à la France. Homme taré, philosophe anarchiste et barbare, les disciples directs de sa morale sont l'ignoble Flachon, et les bandits Bonnot et Garnier. Le gouvernement de la République a décidé de le fêter solennellement au Panthéon, devant le chef de l'État, aux frais du budget national. » Collectif, « Le Métèque Rousseau. Bonnot, Garnier et Flachon au Panthéon », *L'Action française*, 28 juin 1912, p. 1 (italiques dans l'original).

<sup>93</sup> [Anonyme], « La glorification d'un anarchiste », *L'Univers*, 13 juin 1912, p. 1.

<sup>94</sup> L. de Montesquiou, « Rousseau anarchiste », *L'Action française*, 28 juin 1912, p. 1.

<sup>95</sup> F. Mauriac, « En marge d'une plaidoirie pour Jean-Jacques », *Le Figaro*, 19 mai 1942, p. 3.

<sup>96</sup> « Chez nous il a été le père du romantisme. De notre tendance à juger son œuvre à travers les écrivains et la sensibilité romantiques. C'est ce qui explique qu'il est devenu classique en France de parler du sentimentalisme de Rousseau. Sur ce point l'usage est si bien établi qu'on désespère de pouvoir remonter le courant ! » R. Derathé, *Le Rationalisme de Rousseau*, Paris, Presses universitaires de France, p. 181.

ouvrage de jeunesse»<sup>97</sup>, et Jules Lemaître comme «le plus médiocre [de ses] livres [...] celui dont on voit le mieux qu'il aurait pu ne pas l'écrire»<sup>98</sup>. Dans un article de 1909, Faguet tente également de montrer que Rousseau, s'étant rendu compte de ses contradictions en écrivant la première version du *Contrat social*, l'aurait provisoirement abandonné avant d'en reprendre la rédaction en 1762. Il aurait alors tenté d'en supprimer les passages les plus problématiques, «[M]ais les lignes générales du *Contrat social* étaient déjà tracées *ne variarentur* et, en son esprit général, le *Contrat social* est resté ce qu'il était, une pensée de Rousseau contraire à l'ensemble des pensées de Rousseau»<sup>99</sup>. Faguet n'est d'ailleurs pas le seul à voir dans l'œuvre du philosophe de Genève, «deux Rousseau» : celui du *Contrat social* et celui des *Confessions*. Il faudra, en fait, attendre la fin de la seconde guerre mondiale, et les travaux de Robert Derathé notamment, pour que la thèse de l'unité de la pensée de Rousseau s'installe durablement en France<sup>100</sup>.

- 34 Trois observations peuvent être tirées de cette étude. La plus évidente, mais non la moins surprenante, est que Rousseau joue un rôle important dans les débats politiques qui divisent la société française pendant le premier tiers du XX<sup>e</sup> siècle. Ce fait est, je crois, en soi très remarquable. Il semble indéniable, ensuite, que l'utilisation qui en a été faite doit beaucoup à l'interprétation très particulière dont son œuvre a fait l'objet à cette époque, et que cette lecture a été elle-même très influencée par le contexte politique et intellectuel de la période qui l'a immédiatement précédée (les années 1870-1890). Ces résultats, qui mériteraient d'être confirmés sur d'autres époques, confirment la pertinence de la théorie sémiotique interprétative d'Umberto Eco<sup>101</sup> et invitent à l'utiliser plus largement en histoire des idées. Ils nous rappellent, surtout, que les interprétations actuelles de Rousseau, pour l'essentiel inspirées de la tradition kantienne, ne sont ni objectives ni définitives.

<sup>97</sup> E. Faguet, *Rousseau penseur*, ouvr. cité, p. 333.

<sup>98</sup> J. Lemaître, *Jean-Jacques Rousseau*, ouvr. cité, p. 249.

<sup>99</sup> E. Faguet, «La politique de Jean-Jacques Rousseau», *La Revue des Deux Mondes*, 15 septembre 1909, p. 408.

<sup>100</sup> R. Derathé, «L'unité de la pensée de Jean-Jacques Rousseau», dans *Jean-Jacques Rousseau*, S. Baud-Bovy, R. Derathé et al. éd., Neuchâtel, À la Baconnière, 1962, p. 203-218. Certaines études avaient déjà traité du sujet à l'occasion du bicentenaire, en particulier G. Lanson, «L'unité de la pensée de Jean-Jacques Rousseau», *Annales de la Société Jean-Jacques Rousseau*, vol. VIII, 1912, p. 1-32.

<sup>101</sup> U. Eco, *L'œuvre ouverte*, Paris, Le Seuil, 1965.



## AUTONOMIE MORALE ET AUTORITÉ OU LA QUESTION DE LA VOLONTÉ CHEZ ROUSSEAU

Christophe Salvat

L'Harmattan | « Cahiers d'économie Politique »

2007/2 n° 53 | pages 73 à 90

ISSN 0154-8344

ISBN 9782296045774

Article disponible en ligne à l'adresse :

-----  
<https://www.cairn.info/revue-cahiers-d-economie-politique-2007-2-page-73.htm>  
-----

Distribution électronique Cairn.info pour L'Harmattan.

© L'Harmattan. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

# AUTONOMIE MORALE ET AUTORITÉ OU LA QUESTION DE LA VOLONTÉ CHEZ ROUSSEAU

Christophe SALVAT<sup>12</sup>

## Résumé :

La notion de souveraineté renvoie généralement dans la tradition rousseauiste à la souveraineté du corps politique et non à celle des individus comme c'est le cas dans les théories du droit naturel. Rousseau, pourtant, ne cède jamais à la tentation holiste. L'individu, et sa protection, reste au fondement de son système. L'objet de cet article est de questionner le concept d'autonomie morale, en particulier dans sa relation à l'autorité, et de mettre en exergue les effets positifs, notamment en terme de rationalité, de la loi.

## Abstract:

In the Rousseauist tradition, the idea of sovereignty generally refers to the sovereignty of the body politic. Rousseau never yields nevertheless to the holist temptation. Individual is at the core of his system. The aim of this paper is to question the idea of moral autonomy, in particular in its relation to authority, and to emphasize the positive effects of the law on individuals' rationality.

**Classification JEL :** B11 ; B31

## Introduction

La notion de souveraineté renvoie traditionnellement dans les études rousseauistes à la souveraineté du corps politique et non à celle des individus. Rousseau, pourtant, ne cède jamais à la tentation holiste. L'individu, et sa protection, reste au fondement de son système. On connaît ainsi les difficultés théoriques auxquelles se confronte Rousseau à l'égard du statut de la propriété individuelle. L'objet de cet article est de questionner la notion de souveraineté, non du point de vue politique bien connu de tous, mais de celui de l'individu. La notion de volonté individuelle sera notre fil d'Ariane. Elle sera déclinée en trois moments. Dans une première partie, nous mettrons en évidence l'hétéronomie de la volonté et de l'intérêt individuels, en

---

1. CNRS, Greqam ; Robinson College & Centre for History and Economics, Cambridge

2. Je tiens à remercier un rapporteur anonyme de la revue pour ses commentaires. L'orthographe des citations de Rousseau a été modernisée selon les normes de la revue.

distinguant, d'une part, les relations d'autorité exercées par la société, de l'influence du jugement social sur les préférences individuelles, d'autre part. La deuxième partie présente, *a contrario*, le choix individuel tel qu'on peut se le représenter dans sa forme la plus théorique, autonome et rationnelle. Nous nous arrêterons brièvement à cette occasion sur le statut de la fiction dans l'œuvre de Rousseau. Enfin, la troisième et dernière partie s'interrogera sur les conditions de l'autonomie dans l'état civil et sur le rapport conflictuel entretenu entre la volonté individuelle et la volonté générale. Nous essaierons alors de montrer quelle est la nature de la souveraineté du citoyen.

## 1. Intérêt et hétéronomie

Trois formes de dépendance morale (le terme doit être pris dans son sens large i.e. non physique) décrites par Rousseau conduisent l'individu à agir selon un intérêt qui n'est pas à proprement parler le sien, la dépendance morale, la dépendance économique et la dépendance politique. Ces trois types de sujétion de la volonté particulière sont constitutifs de la forme sociale de l'intérêt individuel que Rousseau a désigné par amour-propre. Nous sommes ainsi conduits à distinguer une relation d'autorité qui se traduit par une altération intentionnelle de la volonté d'autrui, qu'elle soit bienveillante ou malveillante, d'une adaptation spontanée de la volonté face au jugement social ou à la représentation que l'individu s'en fait. La première est étroitement associée au statut des inégalités morales dans la société. Plus la société se structure autour d'un rapport à l'inégalité, plus la sujétion de la volonté est nécessaire au maintien de ladite société. Le poids des inégalités morales est renforcé par l'importance accordée au jugement de l'autre dans la construction de l'identité de l'individu. Elle se traduit par une hétéronomie totale de la volonté et l'uniformisation des préférences.

### 1.1. Pouvoir et autorité : la question de la légitimité

Les relations de pouvoir entre personnes forment la première forme de sujétion de la volonté. Elles représentent *a priori* la forme la plus visible de la sujétion car elles impliquent généralement un rapport d'autorité fondé sur l'inégalité. L'idée d'autorité associe au rapport de supériorité, déjà présente dans la notion de pouvoir, l'idée d'une obéissance librement consentie à l'exemple de l'obéissance au père de famille. Elle implique généralement une inégalité naturelle entre celui qui porte l'autorité (le père de famille, Dieu) et ceux qui lui obéissent (les enfants, les croyants). Dans le cas d'une libre sujétion d'un individu envers son semblable, l'autorité naturelle paternelle ou religieuse est instrumentalisée par ces derniers pour asseoir un pouvoir qu'aucun homme, ainsi que l'a justement noté Hobbes, ne serait capable de conserver par la force. A l'instar de l'opinion publique, ces pseudo-relations d'autorité constituent des contraintes, plus ou moins consciemment perçues, mais non moins réelles, de la volonté individuelle. Les deux exemples auxquels Rousseau se réfère sont le paternalisme politique et le dogme religieux.



A la différence de l'autorité légitime de la loi, qui exprime la volonté du corps politique, le pouvoir d'un gouvernement illégitime est une autorité travestie. Elle repose sur la croyance selon laquelle le porteur de l'autorité défend l'intérêt de ceux à qui il commande. La question que soulève la libre sujétion à un pouvoir arbitraire est alors la suivante : Comment parvient-on à tromper les individus sur leur propre intérêt ? Le paternalisme politique constitue une forme historique de tromperie, déjà dénoncée par Locke et Sydney notamment, dans laquelle le gouvernant s'octroie abusivement l'autorité paternelle afin d'agir directement sur la volonté des gouvernés. L'obéissance ne peut être maintenue durablement par la force. Pour être total, le pouvoir doit apparaître comme le fruit même de la volonté de ceux qui y seront assujettis : "l'autorité la plus absolue est celle qui pénètre jusqu'à l'intérieur de l'homme, et ne s'exerce pas moins sur la volonté que sur les actions" (Rousseau 1964a, p.251). Le consentement au faux pacte par les pauvres dans le second *Discours* repose sur un dispositif similaire. Son véritable objet, l'institutionnalisation des inégalités économiques, est l'expression d'une volonté particulière, celle des riches à laquelle les pauvres ne peuvent rationnellement se conformer.

Dire qu'ils sont libres de se déterminer parce qu'ils sont libres de s'exprimer constitue pour Rousseau une contre-vérité. La véritable dépendance se manifeste, au contraire, lorsque les individus expriment la volonté de ceux à qui ils sont assujettis : "Tout l'art du maître, écrit-il dans la *Nouvelle Héloïse*, est de cacher [la] gêne sous le voile du plaisir ou de l'intérêt, en sorte qu'ils [les citoyens] pensent vouloir tout ce qu'on les oblige de faire." (Rousseau 1964b, p.453). Dans le cas d'une autorité légitime, en revanche, l'individu qui obéit suit toujours son véritable intérêt ; il accomplit une volonté qu'il n'est pas capable d'exprimer, mais qu'il peut (s'il en est capable) rationnellement reconnaître comme sienne. La seule autorité légitime naturelle est celle du père. Elle constituera inévitablement un modèle pour l'autorité politique (Rousseau 1964a, p.352). A la différence du paternalisme politique, qui désigne une relation entre un gouvernement et le peuple, l'économie politique légitime caractérise une relation entre le corps politique souverain et les sujets qui en sont membres (Salvat 2006). Le principe fondamental du *Contrat Social* selon lequel "aucun homme n'a une autorité naturelle sur son semblable" (Rousseau 1964a, p.355) est donc respecté dans le cas de l'autorité politique légitime.

A l'instar de l'autorité politique, Rousseau distingue l'autorité morale de la parole divine, révélée par la Bible, parfaitement légitime, de l'autorité arbitraire, ou plus exactement du pouvoir déguisé, de ceux qui s'en réclament (Rousseau 1964a, p.694). La morale constitue une règle de conduite que l'individu, quelle que soit sa religion, se donne à lui-même par la lecture des Écritures ou l'observation de la nature. Le dogme, en revanche, est l'interprétation de la morale que certains hommes, notamment dans la religion catholique, souhaitent imposer aux croyants. En commandant à ses membres un certain nombre de croyances, dont quelques-unes relèvent certes de la Révélation mais d'autres de la superstition voire de la tromperie

(Bernardi & alii 2005), le dogme religieux entretient le culte de l'obéissance et de la crédulité dans l'opinion publique. Ce n'est ni la morale ni la foi des Chrétiens que critique Rousseau dans le *Contrat Social* mais la sujétion qu'elle a instituée. Toutes les religions sont bonnes dès lors qu'elles ne s'octroient aucune autorité sur les hommes ni sur les lois. C'est la raison pour laquelle Rousseau prêche toutes les religions et n'en prêche aucune à la fois: Wolmar est athée, le vicaire savoyard est catholique, Julie est protestante et Émile embrasse la religion naturelle. Chacun trouve dans sa foi une règle de conduite morale qui lui est personnelle, voire intime, mais qui se concilie parfaitement avec la vie civile. Aucune n'est proprement meilleure que l'autre si elle est respectueuse de la libre volonté de chacun. C'est la raison pour laquelle le vicaire savoyard n'est pas prosélyte et le précepteur d'Émile refuse d'aller au-delà de la religion naturelle : "Tant qu'on ne donne rien à l'autorité des hommes ni aux préjugés du pays où l'on est né, les seules lumières de la raison ne peuvent dans l'institution de la nature nous mener plus loin que la religion naturelle, et c'est à quoi je me borne avec mon Émile. S'il en doit avoir une autre, je n'ai plus en cela le droit d'être son guide ; c'est à lui seul de la choisir" (Rousseau 1969, pp.635-36).

## 1.2. Hétéronomie de la volonté et opinion publique

La seconde forme d'hétéronomie de la volonté que nous voulons développer ici n'est ni personnelle ni intentionnelle, elle est sociale. La présence de l'autre, voire la conscience d'une altérité, suffit à modifier nos préférences ou la représentation de notre intérêt. La présence réelle ou virtuelle d'un regard extérieur, tel que le 'spectateur impartial' smithien, n'est jamais légitime car, selon Rousseau, "Sitôt qu'il faut voir par les yeux des autres il faut vouloir par leurs volontés" (Rousseau 1969, p.309). La prise en considération de l'opinion publique dans le processus de choix de l'individu est doublement dommageable. Tout d'abord, elle contrevient à la liberté naturelle de l'homme, à sa capacité de se gouverner soi-même. L'homme trop sensible au regard et au jugement d'autrui n'est plus souverain ; il est moralement dépendant de ses congénères.

A la différence majeure du principe de sympathie, l'opinion n'est ni objective ni moralement structurante. Elle conduit, au contraire, les individus à surévaluer les actions destinées à satisfaire leur *ego* et leur image, au détriment de celles qui leur seraient véritablement utiles et/ou agréables. L'opinion publique représente une forme d'autorité illégitime, bien que diffuse et peu visible, mais dont le rôle institutionnel dans le maintien des inégalités est plus important encore que le gouvernement politique. L'inégalité est, en effet, l'objet même de l'opinion, ce sur quoi se pose le regard du spectateur, et ce sur quoi l'individu espère attirer l'attention. Elle est donc indissociable chez Rousseau de l'idée d'hétéronomie de la volonté.

En l'absence d'un corps politique auquel chacun s'identifie et par lequel chacun est identifié, les individus se construisent une identité sur la différence et la distinc-

tion : "Chacun commença à regarder les autres et à vouloir être regardé soi-même, et l'estime publique eut un prix. Celui qui chantait ou dansait le mieux ; le plus beau, le plus fort, le plus adroit ou le plus éloquent devint le plus considéré, et ce fut là le premier pas vers l'inégalité [...] Sitôt que les hommes eurent commencé à s'apprécier mutuellement et que l'idée de la considération fut formée dans leur esprit, chacun prétendit y avoir droit ; et il ne fut plus possible d'en manquer impunément pour personne" (Rousseau 1964a, pp.169-170).

L'inégalité, en particulier économique, est donc rapidement perçue comme le support d'une valorisation sociale (Salvat 2007). Elle est, comme l'hétéronomie, consentie par ceux-là mêmes qui en sont victimes, elle "s'étend sans peine parmi des âmes ambitieuses et lâches" (Rousseau 1964a, p.188). Or, tous les hommes sont victimes des inégalités, car tous sont également dépendants de l'opinion et des préjugés, les maîtres comme les esclaves, les riches comme les pauvres<sup>3</sup>. En plus d'être trompés par les autres, les individus se mentent à eux-mêmes (*self-deception*) et deviennent dépendants de leurs propres passions. Ils sont, pour reprendre la terminologie de Rousseau, *faibles* ou, plus exactement, la faiblesse de la volonté est l'une des caractéristiques premières de la nature humaine<sup>4</sup>. L'hétéronomie, et le rôle qu'elle occupe dans le maintien et le développement des inégalités, est certes consentie mais elle n'est pas le produit d'un choix raisonné. Elle est la principale manifestation de la faiblesse de la volonté des hommes.

Deux exemples nous permettront de mieux identifier le mécanisme et les effets individuels et sociaux de la sujétion de la volonté individuelle. Le premier, tiré du *Discours sur les arts et les sciences* publié en 1751, présente la sujétion comme une altération des choix, des goûts et des capacités d'un individu, un artiste ou un savant, destinée à satisfaire les normes et les attentes de l'opinion publique. A moins d'abandonner toute vie sociale, celui-ci n'a en effet d'autre choix que de se conformer à la représentation commune de son art. La liberté dont il peut se prévaloir n'est donc qu'illusoire puisqu'elle implique préalablement une normalisation des préférences. Celle-ci n'est généralement pas consciemment perçue. Elle n'apparaît donc pas contraignante pour une très grande majorité d'intellectuels qui ont intégré les normes artistiques ou académiques avant même de se penser comme artistes ou savants. Elle l'est néanmoins pour les rares individus capables de s'élever au-dessus des représentations communes et qui sont par conséquent conscients de leur dépendance. Deux choix s'offrent alors à ce type d'individu, la sujétion ou l'exclusion. Dans le premier cas, tel Voltaire, il "rabaissera son génie au niveau de son siècle",

---

3. "La domination même est servile quand elle tient à l'opinion : car tu dépends des préjugés de ceux que tu gouvernes par les préjugés." (Rousseau 1969, p.308).

4. "C'est la seule tiédeur de notre volonté qui fait toute notre faiblesse, et l'on est toujours fort pour faire ce qu'on veut fortement : *Volenti nihil difficile*." (Rousseau 1969, p.651). Voir aussi la notion de faiblesse dans l'expression "Pourquoi l'homme seul est-il sujet à devenir imbécile ?", (Rousseau 1964a, p.142).

dans le second, il "mourra dans l'indigence et l'oubli" (Rousseau 1964a, p.21). Ces cas demeurent cependant exceptionnels. La quasi-totalité des aspirants artistes ou savants, que Rousseau considère comme 'médiocres', a de fait un intérêt réel à adopter ces normes. En masquant leur inaptitude ou leur faiblesse intellectuelle, elles leur permettent de gagner aussi rapidement qu'indûment un statut social élevé. La sujétion des volontés individuelles par l'opinion publique constitue donc une forme de dépendance insensible et quasiment irréversible, propre à accompagner une oppression politique, et que Rousseau compara à "des guirlandes de fleurs sur les chaînes de fer" (Rousseau 1964a, p.7).

Les normes de consommation constituent une seconde forme de sujétion étudiée par Rousseau. Elles répondent à la fois à la nécessité (ou la faiblesse) morale de se distinguer et à l'impératif économique dicté par la division du travail. Telle qu'elle est décrite dans le *Discours sur l'origine et les fondements de l'inégalité parmi les hommes*, la division du travail structure le lien social autour des besoins économiques. L'échange réduit l'ensemble des inégalités à une seule, plus visible et plus discriminante, la richesse : "Toujours l'objet de l'admiration publique sera celui des vœux des particuliers, et s'il faut être riche pour briller, la passion dominante sera toujours d'être riche" (Rousseau 1964a, p.964). La conjonction de la volonté de distinction et de la volonté de reconnaissance se traduit par conséquent par une uniformisation des préférences. Ces dernières répondent donc à la fois à une exigence économique (satisfaire un besoin ou un désir) et une exigence sociale (se distinguer, se faire admirer).

A la différence de la première, la seconde dimension des préférences individuelles implique la prise en compte du regard et des normes de jugement de ses concitoyens. Elle constitue la part hétéronome de l'intérêt, responsable, selon Rousseau, de la dépendance morale des hommes assemblés en société. La configuration la plus radicale de cette dissonance, la consommation somptuaire, constitue aux yeux de Rousseau un modèle d'irrationalité. En accentuant la dimension relative de son utilité, l'individu se contraint à paraître toujours plus riche au prix de son propre bien-être. Les effets de ces comportements se traduisent au niveau économique par une altération du mécanisme de marché et de détermination des prix (les biens les moins utiles et les plus chers sont prioritairement demandés) ainsi que par une redistribution des facteurs de production vers certaines activités manufacturières ou commerciales au détriment des productions de base telle que l'agriculture (Salvat 2007).

## **2. Autonomie morale et rationalité : rôle des contrefactuels**

L'hétéronomie de la volonté, outre le fait qu'elle questionne un principe de base de la philosophie rousseauiste, celui de la souveraineté, soulève également de nombreuses interrogations sur la rationalité individuelle. L'objet de cette section est d'envisager, d'un pur point de vue théorique, l'autonomie morale de l'individu, c'est-

à-dire un état dans lequel il ne serait sujet d'aucune autre autorité que celle de sa propre raison. Cette représentation contrefactuelle de la souveraineté nous permettra de nous interroger sur les fondements de la notion de volonté, à savoir l'intérêt, ainsi que sur le rôle de la fiction dans l'œuvre de Rousseau.

## 2.1. Intérêt et raison

Les représentations hétéronomes de la volonté et de l'intérêt sont responsables des formes de dépendance exposées précédemment. Elles ne peuvent donc fonder les règles d'un gouvernement libre, que ce soit au niveau de l'individu ou *a fortiori* au niveau du corps politique. Le gouvernement de soi et le gouvernement des autres ont pour principe commun la prévalence de l'intérêt 'naturel' et rationnel<sup>5</sup> (*amour de soi*) sur les motivations sociales et passionnées (*amour-propre*). Le premier est commun à tous les hommes. Son universalité est garantie par sa rationalité. En revanche, en suivant son intérêt tel qu'il peut se le représenter dans le regard des autres (*amour-propre*), l'individu est condamné à une insatisfaction permanente : "L'amour de soi, qui ne regarde qu'à nous, est content quand nos vrais besoins sont satisfaits ; mais l'amour-propre, qui se compare, n'est jamais content et ne saurait l'être, parce que ce sentiment, en nous préférant aux autres, exige aussi que les autres nous préfèrent à eux, ce qui est impossible" (Rousseau 1969, p.493).

L'amour de soi est l'intérêt que chacun peut se reconnaître par une stricte raison et une abstraction totale des éléments sociaux propres à fausser le raisonnement. La volonté autonome de l'individu est, par extension, celle qu'il aurait s'il pouvait rationnellement se déterminer au regard de son seul intérêt (*amour de soi*). La fiction de l'état de nature, ou l'identification à un personnage tel que l'Émile, constituent des expériences de pensée susceptibles de le révéler. Mais, à la différence des Lumières, Rousseau ne prête au raisonnement qu'une faible capacité à assurer l'intérêt commun et individuel des hommes. La faiblesse de la volonté des hommes, entendez par-là leur incapacité à l'autonomie morale, est la principale cause de leur dépendance. L'autorité politique est donc ce par quoi, dans la société civile, l'homme peut être conduit à respecter son intérêt parfois en dépit de ses inclinaisons : "C'est alors seulement que la voix du devoir succédant à l'impulsion physique et le droit à

---

5. Il est important ici de distinguer raison et rationalité. La raison, en tant que faculté, est dissociée du concept de rationalité. Rousseau emploie le terme de raison en lui assignant différentes significations ainsi qu'il l'admet dans *l'Émile* : "Tantôt je dis que les enfants sont incapables de raisonnement, et tantôt je les fais raisonner avec assez de finesse ; je ne crois pas en cela me contredire dans mes idées, mais je ne puis disconvenir que je ne me contredise souvent dans mes expressions." (Rousseau 1969, p.345). Le statut de la raison dans l'état de nature pose en particulier problème si l'on ne distingue pas l'état des facultés cognitives de l'homme et sa capacité à agir dans son meilleur intérêt. Le développement de la raison, en parallèle et en combinaison avec les passions, ne se traduit pas par une rationalité accrue de ses actions. Tout au contraire, plus l'homme raisonne, plus il se compare, plus il veut se distinguer et plus il agit dans ce seul intérêt (*amour-propre*), qui ne correspond cependant pas à ce que l'on pourrait rationnellement attendre de lui (*amour de soi*).

l'appétit, l'homme, qui jusque-là n'avait regardé que lui-même, se voit forcé d'agir sur d'autres principes, et de consulter sa raison avant d'écouter ses penchants." (Rousseau 1964a, p.364).

Les exemples cités ci-dessus expriment, à des niveaux différents, une seule et même idée sur laquelle repose toute la philosophie politique de Rousseau, à savoir que l'indépendance de la volonté individuelle, parfois associée au concept de liberté<sup>6</sup>, n'existe pas en dehors du strict état de nature. L'homme ne peut pas être libre de ses choix s'il considère l'autre comme spectateur de son action et s'il donne à ce jugement une valeur morale ou sociale. Or, et c'est le principal message du second *Discours*, la société civile, à la différence de l'état de nature, se structure spontanément autour de hiérarchies fondées sur des inégalités morales (intellectuelles, économiques, politiques et spirituelles). Plus l'organisation de la société est inégalitaire et plus le regard de l'autre est déterminant dans la détermination des choix individuels.

La relation inverse entre l'importance accordée aux inégalités dans une société et le degré de liberté de ses membres explique la place accordée par Rousseau à l'opinion publique. En comparant et hiérarchisant les individus selon leurs mérites réels ou supposés, elle constitue une véritable caisse de résonance des inégalités sociales et contribue *de facto* à leur développement (Viroli 1988, Shklar 1969). Les lectures les plus critiques de Rousseau voient dans ce processus autoentretenu une impossibilité historique pour l'homme de devenir sujet de sa propre volonté (Descombes 2004). Nous nous proposons de mettre temporairement cette question de côté pour étudier, dans un premier temps, le rapport de l'autonomie morale à la liberté. Si l'on admet que les formes de dépendance présentées ci-dessus constituent aux yeux de Rousseau une antithèse de la liberté, peut-on conclure *a contrario* que la liberté se définit par l'indépendance ? La conception rousseauiste de l'état de nature nous fait rejeter cette hypothèse. A la différence des théoriciens du droit naturel, dont il reprend pourtant la méthode et la terminologie, Rousseau distingue en effet nettement l'indépendance et la liberté<sup>7</sup> (Rousseau 1964a, p.841).

L'état de nature, au sens strict du terme, i.e. un état hypothétique, un contrefactuel, caractérisé par une indépendance totale de l'homme, ne représente ni la des-

---

6. Dans le *Contrat Social*, Rousseau fait une distinction conceptuelle entre trois types de liberté, une liberté naturelle "qui n'a pour bornes que les forces de l'individu", une liberté civile "qui est limitée par la volonté générale" et une liberté morale "qui seule rend l'homme vraiment maître de lui", et qui, précise Rousseau, est un "acquis de l'état civil [...] car l'impulsion du seul appétit est esclavage ; et l'obéissance à la loi qu'on s'est prescrite est liberté" (Rousseau 1964a, p.365). Les deux derniers caractérisent conjointement le concept de liberté développé par Rousseau. Cette distinction est cependant ensuite abandonnée par l'auteur qui se limite à l'emploi du seul terme de liberté. Cela souligne, selon nous, l'indissociabilité de ces deux éléments.

7. Sur la distinction entre sujet autonome et individu indépendant, et sur ses conséquences sur la représentation de la liberté, voir (Descombes 2004, pp.332-339 ; Spitz 1995, pp.368-379).

cription idéalisée d'un âge d'or ni la conjecture d'un passé probable<sup>8</sup> mais une référence théorique négative (ou en creux) de la société civile. L'opération intellectuelle a pour but d'identifier la composante sociale du comportement humain. Rousseau virtuellement 'observe'<sup>9</sup> la nature humaine dans ce contexte totalement désocialisé, puis en déduit la composante acquise. La forte sujétion de la volonté, décrite dans la première section, apparaît mécaniquement comme un effet de la socialisation, mais, et c'est un point fondamental, elle ne signifie pas pour autant que les hommes sont libres dans l'état de nature. L'état de nature rousseauiste soustrait à l'homme toutes ses facultés sociales, dont font partie le langage, la raison et les passions. Seul son instinct le guide dans son action (Rousseau 1964a, p.152). Il n'est donc pas davantage maître de son destin que l'homme civil.

L'autonomie, qui s'oppose ici à la faible volonté, se définit, au contraire, comme la capacité de résister aux passions et opinions sociales lorsque celles-ci contreviennent à notre intérêt autonome (*amour de soi*). La liberté consiste à exercer, par la loi, sa souveraineté en affirmant sa volonté et son intérêt sur ses propres passions. A la différence de l'indépendance, cependant, l'autonomie morale est mal définie. Elle n'appartient proprement ni à l'état de nature ni à l'état civil. La fiction qui, à l'instar d'*Émile ou de l'éducation*, permet de conjuguer les caractéristiques de ces deux états est alors amenée à jouer un rôle prépondérant dans le système rousseauiste.

## **2.2. Le statut de la fiction : Émile**

L'éducation d'Émile, élève explicitement virtuel, est un apprentissage à l'autonomie. Son parcours, de sa naissance à son mariage, a avant tout une portée métaphorique. Émile vit dans un *no man's land* socialement aseptisé. Le genre discursif de cet imposant ouvrage, entre l'essai et le roman, est volontairement difficile à identifier. Il ne répond que très partiellement aux spécificités de la narration et à celles de la fiction. Émile est le fruit de l'imagination de son précepteur, qui lui est bien réel. Il personnifie en quelque sorte l'homme que Rousseau, dans le second *Discours*, s'était figuré dans l'état de nature : "Notre véritable étude est celle de la condition humaine" précise-t-il dans le livre I (Rousseau 1969, p.252). Le choix de la fiction est doublement symbolique. Il traduit, tout d'abord, l'idée que l'imaginaire,

---

8. L'erreur des philosophes issus de la tradition du droit naturel, Rousseau pense précisément à Locke, est de n'avoir pas été assez loin dans l'abstraction, et d'avoir construit à rebours un état de nature à partir de l'état civil (Rousseau 1964a, p.132).

9. L'importance du champ sémantique lié à l'observation dans le second *Discours* a été encore peu commentée à notre connaissance et est souvent malheureusement trahie par ses traducteurs.

à travers notamment l'identification du lecteur à un personnage, peut se révéler être un support de communication plus approprié que l'argumentation<sup>10</sup>.

En jouant sur les écarts qu'elle s'autorise à prendre envers son modèle, la fiction véhicule ensuite un propos critique envers la réalité qu'elle conte<sup>11</sup>. L'autonomie morale à laquelle parvient Émile à la fin de l'ouvrage, au moment même où il s'apprête à devenir père, peut ainsi être interprétée comme une critique de la dépendance dans laquelle les hommes sont tombés, et à laquelle ils peuvent échapper individuellement (Derathé 1962). Mais l'ambition d'*Émile* est, à notre sens, bien supérieure. Le message délivré ici par Rousseau, et que ses *Discours* n'ont pas véritablement été capables d'exprimer, est que l'autonomie morale, cette capacité à se décider rationnellement en-dehors de toute passion sociale, est l'objet même de la loi.

L'autonomie est, tout d'abord, indissociable de l'idée d'autorité. Elle prend, dans le cas d'Émile, deux formes successives, l'autorité de son précepteur, tout d'abord, qui lui apprend à raisonner<sup>12</sup> et l'autorité qu'Émile doit finalement être capable d'exercer sur lui-même. Comme chaque enfant, il a d'abord besoin de l'assistance d'un parent du fait de son incapacité naturelle et temporaire à être autosuffisant. En l'absence d'une autorité véritable, il transforme cependant rapidement cette dépendance en relation de domination. L'autorité doit alors suppléer à l'assistance pour amener l'enfant à se faire dépendant des choses et non des personnes, et lui donner ainsi les outils de sa future autonomie.

Au-delà de l'aspect pédagogique que la question soulève, cette déficience de la raison, présente mais passive, figure d'abord l'incapacité des hommes dans la société civile à se rendre maîtres de leur volonté. Le parcours d'Émile aux côtés de son précepteur<sup>13</sup>, qui du fait de sa fonction a toute autorité sur lui, symbolise le chemin

---

10. On retrouve d'ailleurs cette idée dans le *Contrat Social*, livre II, chapitre VII, dans lequel Rousseau fait reposer l'autorité du législateur sur sa capacité à "persuader sans convaincre" (Rousseau 1964a, p.383).

11. La dénomination de roman, explique Rousseau dans le livre V à propos de l'*Émile*, dépend autant de l'histoire effective des hommes que de celle imaginée de l'un de ses membres : "C'est un assez beau roman que celui de la nature humaine. S'il ne se trouve que dans cet écrit, est-ce ma faute ? Ce devrait être l'histoire de mon espèce : vous qui la dépravez, c'est vous qui faites un roman de mon livre" (Rousseau 1969, p.777).

12. Le but de l'éducation est d'apprendre aux enfants à raisonner et non de raisonner avec eux : "Le chef-d'œuvre d'une bonne éducation est de faire un homme raisonnable, et l'on prétend élever un enfant par la raison !" (Rousseau 1969, p.317). Rousseau reconnaît dans une note du livre II que son vocabulaire peut parfois induire sur ce sujet (comme sur d'autres) à confusion : "Tantôt je dis que les enfants sont incapables de raisonnement, et tantôt je les fais raisonner avec assez de finesse ; je ne crois pas en cela me contredire dans mes idées, mais je ne puis disconvenir que je ne me contredise souvent dans mes expressions." (Rousseau 1969, p.345).

13. Rousseau choisit Émile orphelin dans le livre I, mais précise qu'il "n'importe qu'il ait son père et sa mère. Chargé de leurs devoirs, je succède à tous leurs droits. Il doit honorer ses parents, mais il ne doit



initiatique que les hommes sont appelés à emprunter pour devenir libres. Émile est formé pour vivre parmi les hommes et non en retrait de ceux-ci. Lorsque son éducation prend fin, lorsqu'il s'apprête à devenir père et à rejoindre la société civile, Émile est encore vierge de toute corruption et de toute dépendance : "Jusqu'ici, dit le précepteur à son élève, tu n'étais libre qu'en apparence ; tu n'avais que la liberté précaire d'un esclave à qui l'on n'a rien commandé. Maintenant sois libre en effet ; apprends à devenir ton propre maître" (Rousseau 1969, p.818). L'autonomie morale, ainsi que nous l'avons déjà mentionné, n'est pas un état libre de toute forme de sujétion ou d'interférence. Elle désigne, au contraire, une situation dans laquelle l'individu est capable, généralement par la médiation d'une autorité légitime, d'y résister. Le passage suivant, situé à la fin de l'ouvrage, nous paraît significatif de la portée métaphorique de celui-ci. Après avoir appris à raisonner 'dans le silence des passions', Émile, devenu adulte, reconnaît avoir toujours besoin d'une autorité extérieure pour demeurer maître de lui-même et charge son ancien précepteur d'une nouvelle autorité, la loi, par laquelle il sera capable d'exercer sa propre volonté quelles que soient les circonstances, et non de céder à ses faiblesses :

"Je ne doute pas un instant que si sur ces maximes j'ai su prendre toutes les précautions nécessaires et tenir à mon Émile les discours convenables à la conjoncture où le progrès des ans l'a fait arriver, il ne vienne de lui-même au point où je veux le conduire, qu'il ne se mette même au point où je veux conduire, qu'il ne me dise avec toute la chaleur de son âge, frappé des dangers dont il se voit environné : O mon ami, mon protecteur, mon maître ! reprenez l'autorité que vous voulez déposer au moment qu'il m'importe le plus qu'elle vous reste ; vous ne l'aviez jusqu'ici que par ma faiblesse, vous l'aurez maintenant par ma volonté, et elle m'en sera plus sacrée. Défendez-moi de tous les ennemis qui m'assiègent, et surtout de ceux que je porte avec moi et qui me trahissent ; veillez sur votre ouvrage, afin qu'il demeure digne de vous. Je veux obéir à vos lois, je le veux toujours, c'est ma volonté constante ; si jamais je vous désobéis ce sera malgré moi ; rendez-moi libre en me protégeant contre mes passions qui me font violence ; empêchez-moi d'être leur esclave et forcez-moi d'être mon

---

obéir qu'à moi" (Rousseau 1969, p.267). Dans le livre V cependant, il légitime le choix de sa future femme, Sophie, sur le transfert d'autorité que le père d'Émile a consenti en lui confiant son éducation : "Mon affaire, je dis la mienne et non celle du père, car en me confiant son fils il me cède sa place, il substitue mon droit au sien ; c'est moi qui suis le vrai père d'Émile, c'est moi qui l'ai fait homme" (Rousseau 1969, p.765). Un argument similaire est avancé dans le *Discours sur l'économie politique*. En confiant l'éducation de ses enfants à la nation, le père lui transfère *de facto* son autorité : "Que si l'autorité publique en prenant la place des pères, et en se chargeant de cette importante fonction, l'éducation, acquiert leurs droits en remplissant leurs devoirs, ils ont d'autant moins sujet de s'en plaindre, qu'à cet égard ils ne font que proprement changer de nom, et qu'ils auront en commun, sous le nom de citoyens, la même autorité sur leurs enfants qu'ils exerçaient séparément sous le nom de pères" (Rousseau 1964a, p.260).

propre maître en n'obéissant point à mes sens, mais à ma raison." (Rousseau 1969, pp.651-652).

*Émile, la Nouvelle Héloïse*<sup>14</sup>, ainsi que d'une certaine manière les écrits autobiographiques de Rousseau<sup>15</sup>, expriment une même idée, à savoir que l'autonomie morale est indissociable de la notion d'autorité. Dans le cas d'Émile et de Saint-Preux, elle est personnifiée par un homme de raison, maître de lui-même et de ses passions, et capable de guider son élève ou son ami selon les lumières de la raison. Dans les deux cas, cependant, comme dans le cas de l'éducation d'un enfant, cette tierce autorité n'est que transitoire. L'autonomie n'est jamais assurée, on est toujours tenté par ses passions, la loi prend en quelque sorte le relais de l'éducation.

### 3. Autorité politique et autonomie morale

Dans le contexte politique de l'état civil, la liberté (ou la non sujétion de la volonté) ne peut prendre la forme du retrait ou de l'isolement intellectuel de l'économie domestique de Clarens ou de l'éducation domestique d'Émile. Est libre, dans la tradition républicaine (Pettit 2004, Skinner 2002), celui qui sait rester maître de lui-même face aux aléas du hasard, de la *fortuna* (Pocock 1997), et non celui qui fuit les sources de nuisance. Le citoyen libre est celui qui, dans le même temps, a une totale autorité sur lui-même et qui obéit aux lois. Cet apparent paradoxe est, chacun le sait, l'objet même du *Contrat Social*. Au regard de ce qui a été traité dans la première partie, la difficulté à laquelle se confronte Rousseau concerne la nature et la légitimité de l'autorité de la loi. Elle ne peut être facteur de sujétion, de contrainte ou d'hétéronomie. Nous montrerons donc, dans un premier temps, que le concept de volonté générale est ce par quoi l'individu exerce une autorité sur soi, et que c'est dans le conflit que cela génère qu'il exerce vraiment sa souveraineté.

---

14. La question de l'autonomie morale, et de son apprentissage, prend une nouvelle forme dans la relation de Wolmar et de Saint-Preux. L'ambiguïté du caractère fictionnel de *Julie ou la Nouvelle Héloïse*, quoique plus affirmé que celui d'Émile, est également savamment entretenue par sa forme épistolaire ainsi que sa présentation éditoriale. Le roman s'ouvre sur une situation de dépendance morale, celle de Saint-Preux qui, follement épris de Julie, lui envoie trois lettres successives pour lui signifier son malheur, et la supplier d'y mettre un terme. Il se poursuit comme un apprentissage à la raison et à la maîtrise de soi. Les voyages et l'autorité d'un Mentor constituent les deux éléments essentiels du chemin initiatique de Saint-Preux qui, tel Télémaque, parvient finalement à se rendre maître de ses passions. Wolmar en est, avec Milord Edouard, l'élément clef. Les figures de l'autorité dans la *Nouvelle Héloïse* ont été analysées avec une grande intelligence par Judith Shklar (Shklar 1969).

15. Une étude similaire pourrait être menée à partir de l'expérience même de Rousseau et de ses écrits autobiographiques. Nous avons choisi de ne pas traiter la question ici. Nous renvoyons aux travaux de Starobinski, et notamment (Starobinski 1971).

### 3.1. Avoir autorité sur soi

La nature de la loi, lorsque celle-ci est légitime, est tout d'abord de ne pas être source d'hétéronomie. Dans le cas contraire, elle ne peut être garante de la liberté des individus et manque par conséquent à son principal objet. La loi est une forme d'autorité qui, contrairement aux dites autorités décrites dans la première partie, ne peut influencer ou assujettir les volontés individuelles. Sa nature est définie dans les termes mêmes du contrat social que Rousseau formule ainsi : "Trouver une forme d'association qui défende et protège de toute la force commune la personne et les biens de chaque associé, et par laquelle chacun s'unissant à tous n'obéisse pourtant qu'à lui-même et reste aussi libre qu'auparavant" (Rousseau 1964a, p.360). La forme singulière choisie par Rousseau ('chacun') pour désigner la pluralité des individus est significative. Pour que la loi exprime toutes les volontés particulières, elles doivent logiquement toutes s'identifier à une seule, la volonté générale. Loin de signifier que tous les hommes sont identiques et motivés par un même intérêt, cette formule rend simplement compte de la nature commune et rationnelle de la volonté générale.

La volonté générale est la volonté, non des individus réunis, mais du tout constitué en corps moral : "A l'instant, au lieu de la personne particulière de chaque contractant, cet acte d'association produit un corps moral et collectif composé d'autant de membres que l'assemblée a de voix, lequel reçoit de ce même acte son unité, son *moi* commun, sa vie et sa volonté" (Rousseau 1964a, p.361). La métaphore organique du corps politique, bien que Rousseau lui reconnaisse par ailleurs une pertinence limitée (Rousseau 1964a, p.244), lui permet de construire dans un premier temps une représentation rationnelle et autonome de la volonté. Celle-ci "est toujours droite et tend toujours à l'utilité publique" (Rousseau 1964a, p.371). Le corps politique, à la différence de ses membres, se détermine en fonction d'un intérêt unique et commun à tous. Il est l'intérêt que chacun peut rationnellement et objectivement reconnaître comme sien, celui que l'état de nature a permis d'identifier comme celui de chaque homme (*amour de soi*), mais dont la faiblesse de la volonté prévient en partie l'accomplissement.

Par sa nature même, le corps politique est donc souverain, c'est-à-dire qu'il est capable d'exercer sa libre volonté. Il représente aux yeux de tous l'idée même de la force (Rousseau 1964a, p.360). Mais il n'est qu'une représentation. Le pacte social qui le produit n'est pas un accord entre les individus qui le composent, mais un accord entre soi et l'idée que nous nous faisons du tout. Il s'énonce ainsi : "*Chacun de nous met en commun sa personne et toute sa puissance sous la suprême direction de la volonté générale ; et nous recevons en corps chaque membre comme partie indivisible du tout.*" (Rousseau 1964a, p.361). Le corps politique n'a d'existence que celle que lui reconnaît l'individu. Il prend donc logiquement forme au moment même où l'individu passe un contrat moral avec lui, au moment même où l'individu se reconnaît comme une partie indivisible de celui-ci. Le contrat social associe donc

deux idées, celle d'identité et celle d'autorité. Il se présente d'abord comme une modification par l'individu de sa propre perception parmi les autres. Celui-ci se percevait auparavant comme un *moi* absolu -un individu-, il se perçoit dorénavant comme un *moi* commun -un citoyen- (Rousseau 1969, p.249). Mais, et c'est en cela que le contrat lui est nécessaire, il institue une autorité sur soi-même. Ces deux aspects sont, ainsi qu'on va le voir, indissociables dans la pensée de Rousseau.

Le contrat rousseauiste est, aux dires mêmes de son auteur, particulier dans le fait qu'il ne reconnaît pas la notion d'altérité qui fonde traditionnellement l'idée d'obligation. Chaque individu contracte 'pour ainsi dire' avec lui-même (Rousseau 1964a, p.362). Le contrat est donc un artefact par lequel l'individu s'oblige lui-même. Il se donne en quelque sorte autorité sur lui-même par la médiation de la loi. L'objet de cette convention est, à notre sens, de se donner les moyens d'exercer sa volonté. La faiblesse naturelle de la volonté des hommes les a historiquement conduits vers un état de dépendance et d'inégalité, contraire à l'intérêt de chacun, et qu'aucun homme ne peut rationnellement et objectivement choisir. La nature de la loi est de protéger les hommes, et donc, avant tout, de les protéger contre eux-mêmes. Elle oblige les individus "à conformer leurs volontés à leur raison" (Rousseau 1964a, p.380). La loi ne peut cependant être comparée à une résolution personnelle, un contrat moral que l'on passe avec soi-même, elle tire son autorité de la reconnaissance par l'individu du corps politique et du statut de la citoyenneté qui l'accompagne. En se pensant comme partie d'une communauté, dont tous les membres partagent le même intérêt, l'individu reconnaît dans la règle commune une manière de se conformer à sa propre volonté. Elle "lui apprend à agir selon les maximes de son propre jugement, et à n'être pas en contradiction avec lui-même" (Rousseau 1964a, p.248).

En tant que membres du corps politique souverain, les individus sont libres, c'est-à-dire capables de se gouverner de façon autonome et rationnelle. Les citoyens exercent leur souveraineté par la médiation de la loi. En tant que sujets, cependant, ils sont soumis à une autorité qui contrevient nécessairement à leurs volontés et désirs immédiats : "L'essence de la souveraineté consistant dans la volonté générale, on ne voit point non plus comment on peut s'assurer qu'une volonté particulière sera toujours d'accord avec cette volonté générale. On doit bien plutôt présumer qu'elle y sera souvent contraire ; car l'intérêt privé tend toujours aux préférences et l'intérêt public à l'égalité" (Rousseau 1969, p.843). Il y a de l'aveu même de Rousseau un conflit permanent entre volonté particulière et volonté générale, ou entre intérêt particulier et intérêt commun. L'homme préfère en effet toujours son intérêt à celui de la communauté à laquelle il appartient, et plus la communauté est large plus sa préférence sera accentuée, "En sorte, écrit Rousseau, que chacun est premièrement soi-même, et puis magistrat, et puis citoyen. Gradation directement opposée à celle qu'exige l'ordre social." (Rousseau 1969, p.845). Si la loi rend l'homme libre en lui donnant les moyens d'exercer sa souveraineté, d'agir selon sa volonté autonome, peut-elle le contraindre sans briser la condition d'autonomie, sans devenir elle-

même une source d'hétéronomie ? Si tel est le cas, quel moyen a l'individu de reconnaître une autorité usurpée d'une autorité légitime ? Si la nature de la loi est de forcer l'individu à préférer l'intérêt commun à son intérêt particulier, alors les interprétations faisant de Rousseau une source de légitimité du despotisme sont, croyons nous, irréfutables. L'ambition de notre dernière section est de montrer cependant que ce n'est pas le cas.

### **3.2. La condition de la liberté**

La présence de la loi se concrétise immédiatement pour l'individu par une distance perçue entre sa volonté particulière et la volonté générale. Elle joue donc d'abord un rôle de signalisation. En revendiquant sa légitimité, à savoir le fait, avéré ou non, qu'elle exprime la volonté générale, elle signale la nature de cette distance. L'individu connaît dès lors son intérêt 'objectif'. Il est donc en état de délibérer, à chaque fois qu'une divergence se présente, entre ce que la pression sociale ou ses passions lui dictent, et ce que la raison publique lui prescrit. S'il choisit la première option, il cède à l'hétéronomie et à l'amour-propre; s'il prend la seconde, il choisit l'autonomie et l'amour de soi. Cette action délibérée, faite volontairement sans contrainte ni menace, et répétée à chaque fois que le cas se présente, permet donc à l'individu d'exercer, par la médiation de la loi, une autorité sur lui-même et par conséquent de vaincre la faiblesse de sa volonté.

La distinction entre amour de soi et amour-propre, entre une représentation autonome et hétéronome de l'intérêt particulier, n'est pas historiquement fondée. Pas plus que ne l'est d'ailleurs l'état de nature. Elle établit une distance entre un fait, les hommes sont passionnés et se déterminent comme tels, et un référent, celui de la raison. Malgré la forme faussement narrative adoptée dans le *Discours sur les inégalités*, il n'y a donc pas d'évolution ou de transformation de l'amour de soi en amour-propre, et l'on ne saurait voir *a posteriori* dans l'institution d'un état civil un 'retour' à l'amour de soi. Les volontés particulières et la perception de l'intérêt individuel restent évidemment hétéronomes. L'autonomie morale d'Émile est non seulement fictive mais virtuelle. Rousseau n'attend d'aucun homme une telle perfection. La réflexion politique repose, au contraire, sur les moyens conventionnels de pallier les imperfections naturelles de l'homme ainsi que le rappelle la première phrase du *Contrat social* (Rousseau 1964a, p.351). La divergence entre volonté particulière et volonté générale, évoquée ci-dessus, reflète donc simplement la distance qui existe entre l'amour de soi et l'amour-propre. En exprimant la volonté générale, la loi donne à voir l'intérêt 'naturel' et rationnel partagé par tous. Elle implique l'idée de délibération, de choix voire de conflit entre ce vers quoi nos passions nous conduisent spontanément et ce vers quoi notre raison devrait nous conduire. L'homme raisonnable est donc constamment partagé entre la satisfaction de son intérêt socialisé (*amour-propre*) et celui de son intérêt 'naturel' ou rationnel (*amour de soi*). La nature de la souveraineté réside dans l'exercice de ce choix.

Le point sur lequel nous voulons insister ici, c'est que, loin d'être un état en soi, celui de la sécurité ou de l'absence d'interférence, la liberté se manifeste pour Rousseau dans l'exercice répété de la souveraineté, dans la capacité de l'individu à vaincre la faiblesse de sa volonté. L'état civil légitime garantit certes la liberté politique (la souveraineté du corps politique) mais la liberté civile, c'est-à-dire la souveraineté de l'individu, n'est quant à elle jamais garantie. Elle ne peut l'être, en tout état de cause, lorsque l'individu se croit durablement protégé de toute forme de dépendance (Rousseau 1964a, pp.855-56). La liberté, au sens républicain du terme, se révèle au contraire dans l'épreuve, dans la capacité du sujet à maîtriser la *fortuna* : "il n'y a point de bonheur sans courage ni de vertu sans combat. Le mot de *vertu* vient de *force* ; la force est la base de toute vertu. La vertu n'appartient qu'à un être faible par sa nature et fort par sa volonté (...). Qu'est-ce donc que l'homme vertueux ? C'est celui qui sait vaincre ses affections. Car alors il suit sa raison, sa conscience, il fait son devoir, il se tient dans l'ordre et rien ne peut l'écarter." (Rousseau 1969, pp.817-18).

A l'instar des philosophes républicains, Rousseau présente l'autodétermination à la liberté comme une qualité morale, la force de la volonté ou *vertu*, nécessaire à la citoyenneté (Rousseau 1964a, p.251). Celle-ci ne désigne pas une obéissance absolue et aveugle à la loi mais la reconnaissance, de la part de l'individu, de sa faiblesse naturelle et de son besoin, pour se rendre libre, de la médiation d'une autorité légitime, celle de la loi. La liberté repose sur un équilibre extrêmement fragile, sur un rapport de force toujours incertain entre les volontés particulières et la volonté générale. Le corps politique, explique Rousseau, est destiné à s'autodétruire à plus ou moins longue échéance, il "porte en lui-même les causes de sa destruction" (Rousseau 1964a, p.424). L'objet de la réflexion politique, et de la sienne en particulier, est alors de prolonger cette échéance autant que possible en étant conscient, et en faisant prendre conscience, de la nature humaine et de l'imperfection de toute convention. La souveraineté du corps politique, pas plus que celle de ses membres, n'est jamais acquise. Elle doit, elle aussi, faire l'objet d'une volonté forte et permanente.

## Conclusion

La notion d'autonomie morale que nous avons introduite ici pour aborder la question de la liberté répondait à une source d'insatisfaction. Les distinctions usuelles de liberté positive et négative, ainsi qu'à certains égards leur critique républicaine, nous semblaient particulièrement inappropriées dans le cas de Rousseau dans la mesure où elles réduisent, après Constant, l'indépendance et la souveraineté à la sécurité et la participation citoyenne. Notre ambition a été de montrer combien la question de la volonté individuelle, notamment dans le rapport conflictuel qu'elle entretient avec la loi, sous-tend la pensée de Rousseau. Replacer la volonté de l'individu au centre de son système, orienté ni vers l'indépendance de l'état de nature ni vers une dépendance totale à l'égard de la loi (Spitz 1995), mais vers une autonomie

morale, peut sembler audacieux. Cette lecture permet cependant, pensons-nous, d'interroger différemment l'œuvre de Rousseau, et notamment les notions de faible volonté, de *self-deception* et de souveraineté individuelle. Elle permet également d'appréhender la question de l'unité de la pensée de Rousseau sous un jour nouveau en rejetant, après Durkheim, la perspective historique du second *Discours* et en insistant sur le statut rhétorique de son œuvre fictionnelle. Elle ne prétend pas cependant dans le cadre d'un simple article répondre à toutes les questions ni à combler tous les vides laissés par Rousseau. Le statut de la vertu, rapidement abordé à la fin de cet article, ou celui de la délibération mériteraient, par exemple, d'être plus largement développés. Mais à l'instar de la souveraineté, l'histoire des idées n'est jamais définitive et sa valeur réside essentiellement dans les questionnements et les difficultés conceptuelles qu'elle génère.

## BIBLIOGRAPHIE

- Bernardi Bruno (2005), *La fabrique des concepts. Recherches sur l'invention conceptuelle chez Rousseau*, Paris, Champion.
- Bernardi Bruno, Guénard Florent et Silvestrini Gabriella (2005), *La religion, la liberté, la justice. Un commentaire des Lettres écrites de la montagne de Jean-Jacques Rousseau*, Paris, Vrin.
- Derathé Robert (1950), *Jean-Jacques Rousseau et la science politique de son temps*, Paris, Vrin, 1995.
- Derathé Robert (1962), "L'unité de la pensée de Jean-Jacques Rousseau", in Baud-Bovy S., Derathé R., Dottrens R. et alii, *Jean-Jacques Rousseau, Jean-Jacques Rousseau*, Neuchâtel, A La Baconnière, pp.203-218.
- Descombes Vincent (2004), *Le complément de sujet. Enquête sur le fait d'agir de soi-même*, Paris, Gallimard.
- Pocock John Greville Agard (1997), *Le moment machiavélien. La pensée politique florentine et la tradition républicaine atlantique*, Paris, Presses Universitaires de France.
- Pettit Phillip (2004), *Républicanisme. Une théorie de la liberté et du gouvernement*, Paris, Gallimard.
- Rousseau(1964a), *Du Contrat Social. Ecrits politiques, Œuvres complètes*, vol. III, Paris, Gallimard.
- Rousseau (1964b), *La Nouvelle Héloïse. Théâtre. Essais littéraires, Œuvres complètes*, vol.II, Paris, Gallimard.
- Rousseau (1969), *Emile ou l'Education, Œuvres complètes, vol.IV, Paris, Gallimard*.
- Salvat Christophe (2006), "Oeconomies. Les articles OE/Economie et leurs désignants dans l'Encyclopédie", *Recherches sur Diderot et l'Encyclopédie*, n°40, Octobre, pp.97-116.
- Salvat Christophe (2007), "La loi et l'échange: le statut de la rationalité économique chez Rousseau", *Revue économique*, vol.58, n°2, mars, pp.381-398.
- Shklar Judith (1969), *Men and Citizens. A Study of Rousseau's Social Theory*, Cambridge, Cambridge University Press.
- Spitz Jean-Fabien (1995), *La liberté politique. Essai de généalogie conceptuelle*, Paris, PUF.
- Skinner Quentin (2002), "Un troisième concept de liberté au-delà d'Isaiah Berlin et du libéralisme anglais", *Actuel Marx*, n°32, *Les libéralismes au regard de l'histoire*, pp.15-49.
- Starobinski Jean (1971), *Jean-Jacques Rousseau : La transparence et l'obstacle*, Paris, Gallimard.
- Viroli Maurizio (1988), *Jean-Jacques Rousseau and the 'well-ordered society'*, Cambridge: Cambridge University Press.





## L'ÉCHANGE ET LA LOI

Le statut de la rationalité chez Rousseau

Christophe Salvat

Presses de Sciences Po | « Revue économique »

2007/2 Vol. 58 | pages 381 à 398

ISSN 0035-2764

ISBN 9782724630725

Article disponible en ligne à l'adresse :

-----  
<https://www.cairn.info/revue-economique-2007-2-page-381.htm>  
-----

Distribution électronique Cairn.info pour Presses de Sciences Po.

© Presses de Sciences Po. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

# L'échange et la loi

## Le statut de la rationalité chez Rousseau

---

Christophe Salvat\*

*La pensée économique de Rousseau a souvent été réduite à quelques fragments sur la monnaie, les inégalités, le luxe et l'autarcie. Comparée aux écrits contemporains de Quesnay, Smith ou Hume, elle est donc apparue comme « retardataire », voire réactionnaire. Jean Mathiot a cependant montré que, loin de s'inscrire dans une tradition passéiste, la pensée économique de Rousseau marque au contraire une avancée significative de la philosophie politique dans ses rapports à l'économie politique, dont elle accompagne, voire anticipe, le développement. L'ambition de cet article est de montrer, à travers le cas de l'échange, que la relation entre la politique et l'économie, bien que hiérarchisée, ne se réduit pas à une hégémonie de la rationalité politique. La rationalité économique apparaît, au contraire, constitutive de l'état civil légitime, y compris dans son expression la plus négative, à savoir l'autarcie.*

### ECONOMIC EXCHANGE AND LAW. THE STATUS OF RATIONALITY FOR ROUSSEAU

*The economic thought of Jean-Jacques Rousseau has often been reduced to some pieces on money, inequalities, luxury and autarky. Compared to contemporary writings of Quesnay, Smith or Hume, it has been considered for long as backward. Jean Mathiot has however shown that, far from being backward-looking, Rousseau's economic thought can be seen as a significant move of the political philosophy in its relation with political economy. The ambition of this article is to show, through the example of the economic exchange, that the relation between politics and economics cannot be reduced to an hegemony of the political rationality. The economic rationality appears, on the contrary, constituent of the legitimate civil state including in its most negative expression, I mean autarky.*

Classification JEL : A11, A12, A13, B11, B31, K00.

---

\* CNRS, GREQAM, Centre for History and Economics, Robinson College, Grange Road, CB3 9AN Cambridge. Courriel : cs495@cam.ac.uk

Je remercie les rapporteurs de la revue pour leurs commentaires et suggestions.

## INTRODUCTION<sup>1</sup>

« L'homme est né libre, et par-tout il est dans les fers. » (CS, I, p. 351). Cette célèbre ouverture du *Contrat social* résume presque à elle seule toute la problématique politique de l'auteur. L'homme s'est perdu de lui-même, et son histoire se confond avec celle de son asservissement. Le *Discours sur l'origine et les fondements de l'inégalité parmi les hommes* offre aux lecteurs, après la métaphore de la « guirlande de fleurs sur la chaîne de fer » du *Discours sur les sciences et les arts* publié quelques années auparavant, une véritable déconstruction intellectuelle de la dépendance. Rousseau s'interroge sur sa structure à travers les formes historiques des communautés humaines. Or deux points semblent retenir son attention, le besoin et le lien social. Le besoin est à l'origine de tous les types de communautés (cabanes, tribus, familles, peuples) et le lien social exprime le mode de dépendance (sexuelle, physique, morale, économique, légale). L'échange, économique bien sûr mais plus largement l'échange de regards – la comparaison –, ou l'échange de mots ou de tropes – le langage –, constitue le principal vecteur de la dépendance.

En tentant de comprendre les raisons pour lesquelles l'échange est historiquement devenu le moteur de la dépendance sociale, Rousseau engage une réflexion originale sur la dualité de l'intérêt et l'instrumentalisation sociale et identitaire de celui-ci. Il apparaît alors que la nature de l'échange est étroitement liée à la question de l'hétéronomie (ou de l'autonomie) de l'intérêt, ainsi qu'à la légitimité de l'institution politique. L'autonomie économique, point nodal de l'analyse économique de Rousseau, doit en conséquence être réinterprétée à la lumière de ces rapports. La pensée économique de Rousseau a, en effet, trop souvent été réduite à quelques fragments sur la monnaie, les inégalités, le luxe et l'autarcie. Comparée aux écrits contemporains de Quesnay, Smith ou Hume, elle est donc apparue comme « retardataire », voire réactionnaire. Jean Mathiot a cependant montré que, loin de s'inscrire dans une tradition passéiste, la pensée économique de Rousseau marque au contraire une avancée significative de la philosophie politique dans ses rapports à l'économie politique, dont elle accompagne, voire anticipe, le développement (Mathiot [2000]). Selon lui, la modernité de la pensée économique de Rousseau se révèle paradoxalement dans son ouvrage le plus politiquement achevé, et économiquement le plus épuré, le *Contrat social*.

L'ambition de notre article est de montrer, à travers le cas de l'échange, que la relation entre la politique et l'économie, bien que hiérarchisée, ne se réduit pas à une hégémonie de la rationalité politique. La rationalité économique apparaît, au contraire, constitutive de l'état civil légitime, y compris dans son expression la plus négative, à savoir l'autarcie. Notre approche se fonde pour cela sur un corpus un peu plus large que celui utilisé par Jean Mathiot, et inclut notamment les commentaires de Rousseau sur les écrits de l'abbé de Saint-Pierre

1. Seront utilisées les abréviations suivantes pour les écrits de Rousseau : *CGP* (*Considérations sur le gouvernement de Pologne et sur sa réformation projetée*) ; *CS* (*Du Contrat social ou Principes du droit politique*) ; *DEP* (*Discours sur l'économie politique*) ; *DOI* (*Discours sur l'origine et les fondements de l'inégalité parmi les hommes*) ; *EASP* (*Écrits sur l'abbé de Saint-Pierre*) ; *FP* (*Fragments politiques*) ; *LEM* (*Lettres écrites de la montagne*) ; *MG* (*Du Contrat social ou essai sur la forme de la république – Première version, généralement désigné par Manuscrit de Genève*) ; *PCC* (*Projet de constitution pour la Corse*). Toutes ces références sont tirées de Rousseau [1964b].

ainsi que ses écrits littéraires (*Émile, Julie ou la Nouvelle Héloïse*). La distinction conceptuelle de l'intérêt entre amour de soi et amour-propre constitue le fil conducteur de cette étude. La première partie présente la version critique d'une économie politique (au sens rousseauiste du terme, *i.e.* de gouvernement d'une société) qui s'est structurée autour de la richesse et de sa représentation sociale (Salvat [2006]). L'individuation économique, les inégalités sociales ainsi que l'hétéronomie des intérêts individuels et le dysfonctionnement du marché caractérisent cette économie. Le statut de l'échange dans l'économie politique légitime de Rousseau fait l'objet de la seconde partie. L'autonomie économique y apparaît alors comme la principale limite pratique de la philosophie politique rousseauiste. Elle participe néanmoins dans le discours à instituer une norme de rationalité économique, fondée sur une conception autonome de l'intérêt (*amour de soi*), dont l'institution légitime est tributaire.

## STRUCTURE SOCIALE, INDIVIDUATION ET ÉCHANGE ÉCONOMIQUE

L'intérêt individuel est appréhendé chez Rousseau sous deux formes différentes, l'amour de soi et l'amour-propre<sup>1</sup>. L'un en constitue l'essence ou la norme, l'autre son expression socialisée mais pervertie. L'objet de cette partie est de montrer que cette distinction conceptuelle permet de mettre en évidence la dualité des institutions économiques dans la pensée de Rousseau. Le développement des inégalités, généralement associé à celui du commerce et de la division du travail, a focalisé l'attention des commentateurs sur Rousseau comme critique de l'économie politique<sup>2</sup>. La fiction théorique de l'état de nature a cependant précisément pour but, non de déplorer l'existence perdue d'un âge d'or de l'individu, mais de défendre sa légitimité dans l'état civil. Au niveau économique, cela se traduit par un échange, positif dans son principe, mais dont la nature a été altérée par la dépendance, morale et sociale, de l'individu dans la société.

### La nature de l'échange

La figure rhétorique de l'état de nature est assez commune dans l'Europe du XVIII<sup>e</sup> siècle, et elle est déjà datée à l'époque de Rousseau (Mathiot [2000]). Son emploi est cependant assez singulier chez ce dernier. Il renvoie à la méthode et à la terminologie de la tradition du droit naturel que Rousseau semble, de prime abord, emprunter avant de s'y opposer. Robert Derathé a considérablement contribué à éclairer les relations de Rousseau avec ses prédécesseurs (Derathé [1995]). La question du statut de l'acte économique, dans ses états naturels et

1. « L'Amour de soi-même est un sentiment naturel qui porte tout animal à veiller à sa propre conservation et qui, dirigé dans l'homme par la raison et modifié par la pitié, produit l'humanité et la vertu. L'Amour propre n'est qu'un sentiment relatif, factice, et né dans la société, qui porte chaque individu à faire plus de cas de soi que de tout autre, qui inspire aux hommes tous les maux qu'ils se font mutuellement, et qui est la véritable source de l'honneur. » (*DOI*, n. XV, p. 219.)

2. Voir en particulier, Fridén [1998] et Spector [2003]. Pour une étude plus historique sur les relations entre Rousseau et les physiocrates, voir Bach [1999] et Charles, Steiner [1999].

civils<sup>1</sup>, mérite cependant d'être réexaminée. La théorisation de l'intérêt individuel et de la fonction normale de l'échange dans l'état civil y est étroitement associée.

Il est important, tout d'abord, de rappeler la polysémie du syntagme « état de nature » chez Rousseau. Dans le *Discours sur l'inégalité*, celui-ci peut alternativement renvoyer à l'idée de nature comme à l'état sauvage ou primitif. L'état de nature, à proprement dit, est fictif et hypothétique destiné à « démêler ce qu'il y a d'originaire et d'artificiel dans la Nature actuelle de l'homme » (DOI, p. 123). Il se présente comme un détour analytique, une « expérience de pensée » (Scubla [1992], p. 116), mettant au jour l'irréductibilité de la nature humaine, le « résidu qui ne se laisse pas expliquer par la société » (Burgelin [1973], p. 219). L'état de nature *stricto sensu* est donc par définition a-économique et a-historique. Il se caractérise par une indépendance totale de l'homme. Cette indépendance leur assure, de fait, une liberté naturelle. La satisfaction des besoins est immédiate et illimitée (DOI, p. 135). On ne peut d'ailleurs pas parler à proprement dit de besoins économiques. L'état de nature de l'homme ne comprend donc ni échange ni interaction sociale. Le seul sentiment que les hommes connaissent envers eux-mêmes est la conservation de soi<sup>2</sup> (DOI, p. 140). Le seul penchant qu'ils connaissent envers leurs congénères est la pitié qui « concourt à la conservation mutuelle de toute l'espèce » (DOI, p. 156). Elle implique une forte identification de l'homme à son espèce (DOI, p. 156-157). Le faible développement de leurs facultés d'entendement et le caractère primitif de leur langage assurent spontanément cette identification dans le premier état de nature<sup>3</sup>. La pitié, comme principe d'identification, s'appuie par la suite sur les capacités émotives des individus. Dans un état dans lequel les hommes construisent leur identité personnelle en se différenciant les uns des autres, le sentiment de pitié exige, pour se développer, une faculté de jugement et d'imagination capable de les transporter hors d'eux-mêmes, hors de leurs proches ou de leur famille (EOL, 84 ; Levi-Strauss [1962], p. 243).

La « méthode abstraitive » (Goldschmidt [1983], p. 224) permet à Rousseau de s'inscrire dans la tradition du droit naturel en la poussant dans ses propres retranchements. En montrant que ses prédécesseurs, et notamment Hobbes, se sont arrêtés trop tôt dans leur analyse rétrospective, Rousseau parvient à démontrer que l'homme naturel vit dans un environnement qui lui est favorable, et que, par conséquent, l'état civil ne répond à aucun déterminisme historique (Derathé [1995]). Le second intérêt de cet état fictif est de souligner le caractère historique et contingent du second état de nature, et ainsi de stigmatiser les étapes de la socialisation et de la dépendance. Trois étapes paraissent se dégager. La première se caractérise par l'apparition des notions de besoin et de rareté, la seconde par une structuration positive du groupe autour des inégalités naturelles,

1. Selon Pierre Burgelin, on peut distinguer cinq étapes de la socialisation, l'état de nature, l'état sauvage, l'état barbare, l'état policé et l'état lettré ([1973], p. 272).

2. À quoi ressemble l'homme naturel ? Son langage se résume à un cri sauvage (DOI, p. 148), et sa vie sentimentale à l'instinct sexuel (DOI, p. 158). Ses besoins sont strictement physiologiques, « les seuls biens qu'il connaisse dans l'Univers, sont la nourriture, une femelle, et le repos ; les seuls maux qu'il craigne, sont la douleur, et la faim » (DOI, p. 144). La raison y est totalement absente (DOI, p. 155), voire « contre Nature ». Les hommes sont indépendants les uns des autres, et ne suivent que les pulsions qui leur sont dictées par leur animalité (DOI, p. 143). Les hommes n'ont donc ni l'idée d'eux-mêmes ni des autres. Par conséquent, « n'ayant entre eux aucune sorte de relation morale, ni de devoirs connus, ne pouvoient être ni bons ni méchants, [ils] n'avoient ni vices ni vertus » (DOI, p. 152).

3. En accordant signifiant et signifié, le langage primitif facilite l'identification à l'espèce des individus. F. Markovits [1986], p. 62. Voir aussi J. Starobinski [1971], p. 356-379.

et la troisième par la structuration négative du groupe sous le poids de l'opinion et de l'illusion. C'est dans le cadre de cette troisième période que l'échange économique prend sa véritable dimension sociale. Nous verrons, en seconde partie, que l'état de nature joue enfin un rôle fondamental dans la diffusion d'une norme de l'intérêt, l'amour de soi.

Le second état de nature est une reconstruction historique des premières sociétés. Il permet d'envisager l'origine du lien social dans une perspective économique. Le besoin économique naît de la contingence de ce second état. La Nature, si prodigue dans la première partie du second *Discours*, est ici soumise aux aléas géo-climatiques<sup>1</sup>. Les besoins de l'homme, ou plus exactement la difficulté qu'il a à les satisfaire, la rareté, le contraignent à raisonner, imaginer et communiquer. Les échanges deviennent alors possibles, de même que les engagements mutuels, ainsi que l'illustre l'épisode de la chasse aux cerfs (*DOI*, p. 166-167). Ces interactions restent cependant limitées et réversibles. Aucun rapport de dépendance ou de hiérarchie ne peut naître de ces accords temporels plus ou moins implicites. L'acte économique n'est donc pas à l'origine structurant, et ne constitue pas encore un caractère d'individuation. Mais l'élément déterminant de cette période est sans conteste celui de l'établissement des premières communautés humaines, les familles (*DOI*, p. 167).

La sédentarisation et l'organisation des sociétés autour de la famille marque la seconde étape de la socialisation humaine. L'institution de la famille établit une première faille à l'identification de l'espèce. L'homme appartient d'abord à un groupe familial par lequel il se différencie de ses congénères. C'est, par conséquent, par la cellule familiale que l'individuation économique prend initialement forme. À un second niveau cependant, on observe une volonté d'affirmation et de distinction entre les générations d'une même famille (*DOI*, p. 168). Celle-ci reste cependant limitée car le lien familial, considéré comme naturel, n'implique aucune démarche de reconnaissance du groupe de la part des individus. Les familles se rassemblent les unes autour des autres et « forment enfin dans chaque contrée une Nation particulière, unie de mœurs et de caractères, non par des Réglemens et des Loix, mais par le même genre de vie et d'alimens » (*DOI*, p. 169). En l'absence de constitution politique, le lien social se construit donc sur un mode de vie commun, et non plus sur un intérêt commun comme c'est le cas pour la famille. Cela se traduit par un affaiblissement du sentiment communautaire et par une exigence de reconnaissance beaucoup plus forte de la part des individus : « Chacun commença à regarder les autres et à vouloir être regardé soi-même, et l'estime publique eut un prix. Celui qui chantoit ou dansoit le mieux ; le plus beau, le plus fort, le plus adroit ou le plus éloquent devint le plus considéré. » (*DOI*, p. 169-170.) La société se structure donc, dans un premier temps, autour d'une individuation subjective de ses membres, le plus souvent fondée sur des inégalités naturelles. L'inégalité des talents se traduit nécessairement par une inégalité productive, « le plus fort faisoit plus d'ouvrage ; le plus adroit tiroit meilleur parti du sien ; le plus ingénieux trouvoit des moyens d'abrèger le

1. La variation des saisons, des climats, des terrains, ou les années stériles, les accidents climatiques ou naturels (volcans, tremblements de terre) menacent l'approvisionnement des hommes. Tout semble rendre la vie humaine difficile, « la hauteur des Arbres, qui l'empêchoit d'atteindre à leurs fruits, la concurrence des animaux qui cherchoient à s'en nourrir, la férocité de ceux qui en vouloient à sa propre vie, tout l'obligea de s'appliquer aux exercices du corps ; il falut se rendre agile, vite à la course, vigoureux au combat » (*DOI*, p. 165).

travail » (*DOI*, p. 174). Elle fait de la reconnaissance sociale des individus, fondée sur la distinction et la supériorité, son principal vecteur d'intégration sociale.

Dans l'idéal rousseauiste, le rapport d'échange devrait permettre de hiérarchiser les activités économiques en fonction de la reconnaissance sociale qui leur est due (*DOI*, n. XIX, p. 223)<sup>1</sup>. Celle-ci est conjointement définie par l'utilité sociale de la production, le fer doit avoir plus de valeur que l'or (Rousseau [1969], p. 295), et par sa faible corrélation aux autres productions<sup>2</sup>. On trouve donc naturellement au sommet de l'échelle de valeurs de Rousseau l'agriculture, puis la forge, puis la charpente (Rousseau [1969], p. 296). En revanche, si les différences de fortune sont inévitables, elles ne doivent jamais constituer une source d'autorité morale. Les inégalités économiques ne sont injustes pour Rousseau que lorsqu'elles ne correspondent pas à une inégalité naturelle (*DOI*, p. 193-194). L'égalité de fortune n'est d'ailleurs ni possible ni même souhaitable (*FP*, VII, p. 522). On voit par là que l'échange n'est pas théoriquement condamné par Rousseau, et qu'il peut même, au contraire, se révéler être un levier utile dans un état légitime. Il présente en effet trois atouts essentiels à cette fonction, il peut mesurer, valoriser et hiérarchiser les inégalités naturelles. Le *Discours sur l'inégalité* associe cependant étroitement l'échange au développement des inégalités morales, de l'injustice et de la dépendance.

## Amour-propre et dénaturation de l'échange

Comment passe-t-on dans l'analyse de Rousseau d'un échange, dans un premier temps (logique mais non historique) facteur d'individuation positive, à un échange responsable de la structuration sociale autour des inégalités morales ? Dans la conjonction d'éléments qui caractérisent la fin de l'état de nature, on peut distinguer trois points : la perte du référent identitaire, l'instrumentalisation sociale de la fonction d'échange et l'hétéronomie croissante de l'intérêt individuel. Ces altérations du rapport de l'homme à son environnement sont toutes trois comprises dans ce que Rousseau a nommé *amour-propre*, par opposition à l'*amour de soi* de l'homme naturel. L'un et l'autre expriment ce que l'on désigne communément par intérêt. L'amour de soi est la forme naturelle de l'intérêt dans laquelle « chaque homme en particulier se regardant lui-même comme le seul Spectateur qui l'observe, comme le seul être dans l'univers qui prenne intérêt à lui, comme le seul juge de son propre mérite » (*DOI*, n. XV, p. 219) se détermine en fonction de sa seule utilité. L'amour-propre est la forme sociale que prend l'intérêt individuel. Il « n'est qu'un sentiment relatif, factice, et né dans la société, qui porte chaque individu à faire plus de cas de soi que de tout autre, qui inspire aux hommes tous les maux qu'ils se font mutuellement, et qui est la véritable source de l'honneur »

1. La hiérarchisation sociale des individus, en fonction de leur capacités respectives et lorsqu'elle est pleinement acceptée par chacun d'eux, constitue ce que M. Viroli désigne par « société bien ordonnée ». En prenant exemple sur la *Nouvelle Héloïse*, il montre que l'ordre de la société est indissociable chez Rousseau de la hiérarchie. Sur la relation entre « ordre », « désordre », et « inégalités », voir en particulier son chapitre 3 « Disorder and inequality », Viroli [1988], p. 53-106.

2. « Il y a un ordre non moins naturel et plus judicieux encore par lequel on considère les arts selon les rapports de nécessité qui les lient, mettant au premier rang les plus indépendans, et au dernier ceux qui dépendent d'un plus grand nombre d'autres. Cet ordre qui fournit d'importantes considérations sur celui de la société générale est semblable au précédent et soumis au même renversement dans l'estime des hommes. » (Rousseau [1969], p. 295.)



(DOI, n. XV, p. 219). Il est, selon Rousseau, responsable de la dénaturation des principales institutions humaines, comme l'échange mais aussi le langage, les arts ou même les lois. Rétablir ces institutions dans leur fonction originelle nécessite un retour de l'homme sur ses déterminants naturels (besoins absolus, expression sincère des sentiments, notions du beau et du juste véritables).

En l'absence de loi, l'autorité se fonde sur les différences individuelles, qui sont de quatre types (richesse, noblesse, puissance et mérite), mais qui toutes se réduisent *in fine* à la seule inégalité économique « parce qu'étant la plus immédiatement utile au bien-être et la plus facile à communiquer, on s'en sert aisément pour acheter tout le reste » (DOI, p. 189). Le choix d'un critère d'individuation qui soit relatif, la richesse (FP, VII, p. 521), est un facteur important de compétition et de domination sociale. Il implique une forte rivalité entre les membres du groupe, « l'ambition dévorante, l'ardeur d'élever sa fortune relative, moins par un véritable besoin que pour se mettre au-dessus des autres, inspire à tous les hommes un noir penchant à se nuire mutuellement, une jalousie secrète d'autant plus dangereuse que, pour faire son coup plus en sûreté, elle prend souvent le masque de la bienveillance » (DOI, p. 175). Le marché, en l'absence d'un cadre légal légitime, peut donc se comparer à un état de nature, au sens hobbesien du terme, pour lequel la force constitue la seule source de légitimité. L'objet de l'échange n'est à proprement parler ni réel ni monétaire, il est moral. La satisfaction du bien acquis est à la hauteur de la privation qu'elle représente pour les autres car « ce n'est qu'en raison des différences que les avantages de la fortune se font sentir » (FP, VII, p. 522). Le riche, écrit encore Rousseau, cesserait d'être heureux si le peuple cessait d'être misérable (DOI, p. 189). L'échange est donc nécessairement producteur d'inégalité morale lorsqu'il est source de lien social. En structurant les rapports sociaux, l'échange concurrence tout autre principe d'organisation civile, notamment politique. À terme, ce système social fondé sur l'inégalité morale ne peut que se traduire par un retour à un nouvel état de nature, où seule la force fait loi, où les notions même de bien et de justice ont disparu et où « tous les particuliers redeviennent égaux parce qu'ils ne sont rien<sup>1</sup> » (DOI, p. 191).

L'importance de la question identitaire dans le processus de socialisation est fondamentale, ainsi que l'a bien montré Maurizio Viroli (Viroli [1988], p. 96). L'amour-propre est, en effet, fortement lié à une quête de reconnaissance sociale des individus. Il se définit paradoxalement comme le désir individuel de se distinguer du groupe, de s'en faire admirer, pour en être reconnu. La sédentarisation et le rassemblement des familles, autrement dit la constitution de tribus, en créant une nouvelle communauté, déplace le référent identitaire vers la communauté dominante. Le lien naturel qui unit l'individu à la famille n'est plus suffisant lorsque les familles se regroupent entre elles. Or, en l'absence de constitution d'un

1. La fin du *Discours sur l'inégalité* s'achève sur cet étrange retour aux origines, un état de nature qui n'est plus l'état de paix et d'indépendance que Rousseau s'était plu à décrire, mais bien l'état d'anarchie et de guerre décrit par Hobbes et dont il contestait la version. La forme circulaire adoptée par Rousseau dans le second *Discours*, mais on pourrait également étendre l'analyse au *Discours sur les sciences et les arts*, s'inspire du schéma rhétorique de la fable dont on retrouve parfois les figures stylistiques (DOI, p. 192). Or derrière la fable se profile traditionnellement une morale, parfois explicite, mais également une forte critique politique. La morale qui clôt le *Discours* moque l'autorité morale de la richesse au travers d'une analogie stigmatisant le grotesque et le monstrueux d'une société inégalitaire (DOI, p. 193-194). Sur la bonne utilisation des fables dans l'éducation, voir Rousseau [1969], p. 188, Voir aussi J. Starobinski, « Fable et mythologie aux XVII<sup>e</sup> et XVIII<sup>e</sup> siècles », *Critique*, n° 366, 1977 ; repris dans Starobinski [1989], p. 233-262.

corps politique, l'appartenance des individus au groupe ne repose plus que sur l'opinion. La quête d'une identité sociale, et non plus simplement domestique, fait de la reconnaissance sociale une véritable obligation, « chacun prétendit y avoir droit ; et il ne fut plus possible d'en manquer impunément pour personne » (*DOI*, p. 170). Le rapport de l'homme à lui-même, jusque-là circonscrit à ses seuls besoins (*amour de soi*), est alors soumis à la représentation qu'il se fait de lui-même au regard des autres. Il vit, selon les termes mêmes de Rousseau, « en dehors de lui-même ». L'échange joue un rôle de médiation ou d'interface entre l'individu et la communauté à laquelle il aspire.

Le lien social, dans sa forme spontanée, est donc premièrement économique. Il légitime d'abord la réunion des individus par l'impératif de conservation, puis en hiérarchise l'organisation autour de l'échange. Cela se traduit par l'altération de l'échange lui-même, notamment par la nouvelle symbolique de la monnaie, mais également des intérêts et des besoins des individus<sup>1</sup>. L'intérêt économique répond, en effet, dans cette configuration, à deux rationalités différentes, la conservation et l'identification. Cela a deux conséquences majeures sur l'échange. La première est que les individus poursuivent un intérêt dicté par des considérations extérieures. L'intérêt des agents est réorienté vers l'acquisition, non comme moyen de satisfaire ses besoins (*amour de soi*), mais comme moyen de se valoriser (*amour-propre*). L'échange ne répond par conséquent plus à une rationalité économique pure. Les biens de luxe, dont le prix fait la vraie valeur, illustrent bien avec les mouvements de mode la nouvelle rationalité engendrée par une société du paraître. Ils ne sont désirés que pour leur forte visibilité sociale et leur caractère discriminant. Ils produisent immédiatement la reconnaissance sociale : « Toujours l'objet de l'admiration publique sera celui des vœux des particuliers, et s'il faut être riche pour briller, la passion dominante sera toujours d'être riche. » (*CGP*, p. 964).

L'hétéronomie des intérêts, outre le fait d'être psychologiquement et socialement destructrice, pervertit également les principes mêmes de l'échange, notamment ceux de la détermination des prix et de la division du travail. Les prix des biens, selon Rousseau, se fixent en raison inverse de leur utilité (*DOI*, n. IX, p. 206). Plus les biens demandés sont superflus, plus le rapport d'échange devient défavorable aux agriculteurs. Le développement des métiers de luxe se traduit donc inévitablement par un abandon des campagnes et des métiers les plus utiles socialement. La seconde conséquence est relative à la nature de la monnaie. La monnaie est un signe représentatif des biens. C'est un bien choisi pour représenter les autres dans l'échange, une simple mesure des autres biens<sup>2</sup>. Lorsque la richesse devient la « passion dominante » de la société, la monnaie est alors désirée pour elle-même, pour le symbole de réussite et de puissance qu'elle offre (*FP*, VII, p. 522), « d'où il suit qu'un État peut se trouver dans une telle situation qu'avec une fort grande quantité d'argent il ne laisseroit pas d'être réellement très pauvre et de manquer du nécessaire » (*FP*, VII, p. 521). Enfin, le caractère relatif de la richesse, et de sa composante monétaire, accentue sensi-

1. Il existe, selon M. Viroli, trois types de besoins chez Rousseau, le besoin de subsistance, de bien-être et le besoin moral que l'individu détermine en fonction de l'opinion. Ce dernier est un épiphénomène de l'amour-propre, M. Viroli, [1988], p. 97.

2. Sa valeur est déterminée par la valeur des métaux qui la composent. Rousseau condamne toute tentative du législateur de vouloir en fixer ou en altérer la valeur. Il a une bonne connaissance des mécanismes monétaires, et attribue les variations de la valeur réelle de la monnaie à son abondance ou sa rareté, l'abondance ou la rareté des biens, et sa vitesse de circulation, elle-même déterminée par la quantité des échanges effectués (*FP*, VII, p. 520-521).

blement la rivalité dans l'échange et fait de l'inégalité, non plus simplement une conséquence presque nécessaire de celui-ci, mais l'objet même de l'échange.

La monnaie, indissociable de sa représentation sociale, se pose donc comme un second référent fort embarrassant. Elle nuit, tout d'abord, au développement de l'agriculture au profit des activités financières et commerciales plus lucratives mais moins utiles (*PCC*, p. 920 ; *CGP*, p. 1008). Le commerce, en particulier, attire un nombre croissant d'individus par les profits importants qu'il peut générer. Celui des produits agricoles, pourtant nécessaire, n'échappe pas à la règle, il décourage plus qu'il ne suscite le développement de l'agriculture et des arts utiles (*PCC*, p. 920). Le goût de la richesse, ensuite, ne stimule pas le travail mais est, au contraire, source de paresse et de vanité (Rousseau [1969], p. 359). L'établissement de régies locales, rassemblant les offres et les demandes de chacun et établissant des rapports d'échange, est évoqué par Rousseau pour réduire l'empire de la monnaie sur les échanges (*PCC*, p. 923-924). Ces échanges pourraient également s'effectuer par l'intermédiaire d'une monnaie fictive : « Ces opérations peuvent se faire avec la plus grande justesse et sans monnaie réelle, soit par la voie d'échanges ou à l'aide d'une simple monnaie idéale qui servirait de terme de comparaison telle par exemple que sont les pistoles en France, soit en prenant pour monnaie quelque bien réel. » (*PCC*, p. 923.) De plus, la monnaie n'est que trop utilisée par l'administration elle-même, notamment par le biais de la fiscalité et des dépenses publiques. Il est donc nécessaire d'en rationaliser l'usage. Rousseau propose donc l'établissement d'un domaine foncier public (*DEP*, p. 264-265 ; *PCC*, p. 931)<sup>1</sup>. Les dépenses publiques monétaires doivent, quant à elles, être réduites autant que possible. La finance publique est un des points névralgiques de l'analyse économique de Rousseau. Ce dernier la compare au « sang qu'une sage économie, en faisant les fonctions du cœur, renvoie distribuer par tout le corps la nourriture et la vie » (*DEP*, p. 244). La distinction entre finance publique et activité monétaire est fondamentale chez Rousseau. La finance publique est essentielle au bon fonctionnement économique de la Nation, mais elle ne doit pas se réduire à son caractère monétaire. La finance publique est « le nerf de la force publique » (*PCC*, p. 930), mais l'usage de la monnaie l'affaiblit. Rousseau la compare à la graisse véhiculée par le sang (*PCC*, p. 931). En limiter la circulation ne signifie donc pas limiter l'action publique, mais la gérer plus rationnellement (*CGP*, p. 1007-1008). La monnaie, comme le langage, représente un frein à la dénaturation de l'homme et à l'établissement d'une vraie société civile. Son usage doit donc être aussi limité que possible à l'intérieur comme à l'extérieur de la nation.

## AUTONOMIE, LOI ET ÉCHANGE

La valorisation sociale et individuelle des richesses, loin d'être source de progrès, contrarie le développement normal de l'économie nationale. Rousseau se démarque donc sensiblement de ses contemporains tels que Smith ou Hume. La richesse et la loi constituent deux référents exclusifs l'un à l'autre<sup>2</sup>. Il n'est

1. Voir aussi Bouvignies [2002].

2. « Jamais, écrit Rousseau dans la *Lettre à d'Alembert*, dans une monarchie l'opulence d'un particulier ne peut le mettre au-dessus du prince ; mais dans une République elle peut aisément le mettre au-dessus des lois. Alors le gouvernement n'a plus de force, et le riche est toujours le vrai souverain. » (Rousseau [1967], p. 171.)

donc pas possible d'avoir dans le même temps deux sources d'autorité, la loi et la richesse. Or, seule la première est légitime<sup>1</sup>. Quelle est alors la place de l'échange dans l'état civil ? Nous montrerons que la loi assure, en quelque sorte, la rationalité de l'échange en limitant la pratique aux usages conformes à l'amour de soi. La seconde partie de cette section s'interroge sur le statut paradoxal de l'autarcie dans l'œuvre de Rousseau et notamment dans ses fictions. Le rôle du philosophe, à l'instar du législateur, serait donc de substituer aux représentations collectives hétéronomes une nouvelle norme de rationalité. Cet apprentissage passe par une représentation autonome, voire autarcique de l'individu ou de la communauté, et renvoie, au-delà de l'une des figures traditionnelles de l'enseignement économique, Robinson Cruséo, au statut de la fiction dans la conceptualisation de la rationalité économique.

## L'échange dans l'économie politique légitime

L'autorité morale de l'opinion ne permet pas aux hommes d'agir en fonction de leur véritable intérêt (*amour de soi*). En se soumettant à une autorité morale économique, sensée justement représenter les inégalités naturelles, les individus se déterminent en fonction d'un impératif de reconnaissance et de pouvoir que leur assure par ailleurs la richesse (*amour-propre*). L'état policé, décrit dans le premier puis second *Discours*, et dans lequel les individus agissent dans une apparente indépendance, ne doit sa stabilité qu'à l'hypocrisie et aux normes de civilité de ladite civilisation (Starobinski [1989], p. 165 et *sq.*). Le passage de l'état de nature à l'état civil induit, en revanche, un transfert de l'autorité morale de la richesse vers la loi. Celle-ci assure un statut identique à tous les citoyens. Les hommes, qui, dans l'état de nature, sont tous égaux devant la Nature, sont à l'état civil tous égaux devant la loi : « Chacun de nous met en commun sa personne et toute sa puissance sous la suprême direction de la volonté générale ; et nous recevons en corps chaque membre comme partie indivisible du tout. » (CS, I, p. 361.) Cette égalité réinventée donne à chacun une identité commune qui transcende les identités personnelles forgées dans la différence (Viroli [1988], p. 191). Dépendants de la nature dans l'état de nature, les hommes restaurent leur liberté en s'extrayant des dépendances personnelles pour se soumettre collectivement et absolument à la loi (Spitz [1995], p. 368-379). C'est donc paradoxalement dans la réification et la sacralisation de la loi que dépend la liberté civile des hommes (Scubla [1992])<sup>2</sup>. Le législateur représente l'autorité capable d'établir la conformité des intérêts particuliers à l'intérêt de la nation, de faire « régner la vertu » (DEP, p. 252). La liberté civile, en abolissant le système de dépendance arbitraire auquel étaient soumis les hommes, leur permet donc de retrouver une forme d'autodétermination et de rationalité<sup>3</sup>.

---

1. « La puissance civile s'exerce de deux manières : l'une légitime par l'autorité, l'autre abusive par les richesses. » (PCC, p. 939.)

2. Son analyse est, sur ce point, inspirée des travaux de René Girard.

3. Dans un contexte civil, la liberté ne peut être associée à la notion d'indépendance, ainsi que le précise Rousseau lui-même dans ses *Lettres écrites de la Montagne* (LEM, VIII, p. 841). La vraie liberté ne peut se résumer à une absence factuelle de contrainte, elle réside dans la garantie de l'absence de contrainte. La tradition républicaine ne s'oppose pas à la notion de liberté négative, elle compose avec. Philip Pettit et Quentin Skinner ont largement contribué ces dernières années à défendre le concept de liberté républicaine, entendu comme une non domination (par opposition à la non-coercition de la liberté négative). P. Pettit [2004]. Voir aussi Q. Skinner [2002].

Que devient l'échange dans l'état civil légitime ? Ou, plus exactement, quel rôle est-il amené à jouer dans une économie dans laquelle la loi représente la seule autorité morale légitime ? Rousseau ne s'est pratiquement pas exprimé sur le sujet, même si, on le sait, la seconde partie des *Institutions politiques*, dont est issu le *Contrat social*, et qui n'a jamais été rédigée, devait traiter du commerce. Un bref passage de l'*Émile* consacré à la monnaie vient cependant rappeler la neutralité de son principe : « Nulle société ne peut exister sans échange, nul échange sans mesure commune, et nulle mesure commune sans égalité. Ainsi toute société a pour première loi quelque égalité conventionnelle soit dans les hommes soit dans les choses. L'égalité conventionnelle entre les hommes, bien différente de l'égalité naturelle, rend nécessaire le droit positif, c'est-à-dire le gouvernement et les lois. [...] L'égalité conventionnelle entre les choses a fait inventer la monnaie ; car la monnaie n'est qu'un terme de comparaison pour la valeur des choses de différentes espèces, et en ce sens la monnaie est le vrai lien de la société. » (Rousseau [1969], p. 297.) Dans un état légitime et vertueux, la médiation de l'échange n'est plus, en théorie, nécessaire à l'institution d'un lien social. L'état civil, en assurant par la loi, l'identité collective et la liberté des individus, permet donc théoriquement de restaurer la rationalité et la légitimité de l'échange. En devient-il pour autant libre ?

Trois considérations nous amènent à penser que la rationalité économique ne se conjugue pas nécessairement pour Rousseau avec la liberté de l'échange. Tout d'abord, il est important de rappeler que, contrairement à l'amour-propre, l'amour de soi est une abstraction. Il représente une norme rationnelle de l'intérêt individuel, mais n'a pas de dimension réelle ou historique. Le contrat social s'appuie, comme la loi, sur une conception autonome de l'intérêt (*amour de soi*) mais il ne signifie pas que les individus agissent comme tel. Leurs motivations restent socialement déterminées. Rousseau ne rappelle-t-il pas, dès la première phrase du *Contrat social*, qu'il prend « les hommes tels qu'ils sont, et les lois telles qu'elles peuvent être » (CS, p. 351) ? Les motivations de l'échange ne sont donc pas modifiées par la loi, seules ses pratiques peuvent l'être. En conséquence, la loi est légitime lorsqu'elle encadre l'échange dans les limites que lui impose l'intérêt. Elle l'interdit ou le contraint lorsque les individus, emportés par l'envie et la vanité, cèdent à des désirs qui vont manifestement à l'encontre de l'intérêt commun et de leur intérêt particulier. C'est le cas notamment du commerce du luxe. La légitimité de l'échange dépend donc de sa rationalité économique, elle-même assurée par la loi. La liberté de l'échange est garantie par la loi en même temps que l'autonomie morale des citoyens. Il y a donc, pour reprendre, sur ce point l'analyse de Jean Mathiot, une prééminence de la rationalité politique sur l'économique. L'architectonique de la politique, selon son expression, désigne conjointement la capacité supérieure de la loi sur la richesse à instituer un ordre social, ainsi que sa supériorité heuristique. Non seulement la solution la plus juste est la plus appropriée, explique Jean Mathiot, mais elle peut également s'exprimer en terme d'intérêt (Mathiot [2000], p. 28).

L'existence de peuples historiquement constitués représente, ensuite, un véritable frein à l'institution d'une nouvelle norme de rationalité. Leurs lois et leurs coutumes, bien que souvent injustes et illégitimes, ont été acceptées et enseignées de générations en générations. Elles ont *de facto* participé à la formation d'une représentation collective de l'intérêt (*amour-propre*), lui-même légitimé par certains philosophes, Mandeville notamment. Vouloir réformer ces croyances est une entreprise, non seulement difficile, mais parfois aussi vaine et dangereuse (CS,

II, p. 385). Le peuple le plus propre à la législation est donc celui qui « n'a point encore porté le vrai joug des loix ; celui qui n'a ni coutumes ni superstitions bien enracinées » (CS, II, p. 390). La Corse semble être l'un des seuls pays à répondre à l'ensemble des conditions posées par Rousseau<sup>1</sup>. Mais même, dans ce dernier cas, l'établissement d'une véritable citoyenneté est compromis par le poids des représentations morales de l'échange. Les préjugés des pays riches y sont en effet déjà répandus, « préjugés qu'il faut combattre et détruire pour former un bon établissement » (PCC, p. 902). Les hommes sont égaux de droit, mais restent inégaux dans l'opinion. Les représentations collectives de l'échange restent donc empruntées d'une symbolique passiviste. Des mesures de restriction, comme la limitation de la circulation de la monnaie, semblent donc d'abord répondre à des mesures de prudence et de pragmatisme politique<sup>2</sup>. Elles participent également d'une volonté de réforme des représentations individuelles et collectives de l'intérêt.

Enfin, l'influence de la loi n'est, par définition, possible que dans le cadre de la nation. Les échanges internationaux y échappent nécessairement. Dotés d'une personnalité morale, les corps politiques sont, en effet, eux-mêmes soumis au joug des passions dont ils sont censés préserver leurs membres. Au niveau international, les nations sont donc encore dans un état de nature. L'échange constitue d'abord, comme la guerre, un mode d'affirmation de pouvoir et d'identité des nations<sup>3</sup>. Cela se traduit, à l'instar des individus, par une distorsion de l'utilité et du rapport d'échange. L'état hybride qui caractérise alors les sociétés est donc pour Rousseau le pire de tous<sup>4</sup>, car « vivant à la fois dans l'ordre social et dans l'état de nature, nous sommes assujettis aux inconvénients de l'un et de l'autre, sans trouver la sûreté dans aucun des deux » (EASP, p. 610). Les relations entre nations s'inscrivent dans une démarche collective (et non plus individuelle), conflictuelle et identitaire. L'échange international répond alors aux exigences de visibilité et de discrimination de l'amour-propre. L'agriculture et les productions de première nécessité sont dévalorisées, et *in fine* délaissés, au profit d'activités plus lucratives telles que le commerce du luxe. Laisser l'échange libre entre les nations constitue, par conséquent, une menace pour la conservation du corps politique et de ses membres. Or, en l'absence d'une communauté européenne, voire mondiale, l'autonomie morale ne peut être réalisée que dans le cadre d'une communauté réduite et elle-même autonome. Une liberté totale de l'échange, pour répondre à l'exigence d'harmonisation des intérêts, impliquerait au niveau politique un pacte social universel.

1. Outre l'absence de lois, l'indépendance, l'état de paix, la taille et l'absence de richesses sont des éléments fondamentaux pour l'établissement d'un peuple nouveau (CS, II, p. 391).

2. L'économie publique ou politique s'inscrit chez Rousseau dans la tradition de l'économie entendue comme *prudens administratio* à l'image du gouvernement sage et prudent du père de famille. La prudence est une propriété longtemps associée à l'idée même d'économie, et qui trouve probablement ses racines chez Aristote pour lequel elle constitue une vertu première. Voir P. Aubenque [1963]. Voir aussi sur la sémantique de l'économie dans l'*Encyclopédie*, et notamment dans l'article de Rousseau, Salvat [2006].

3. Dans ses commentaires sur les écrits de l'abbé de Saint-Pierre, Rousseau envisage cependant l'hypothèse d'une fédération des nations européennes seule capable de concilier la notion communautaire de l'amour de soi et l'enjeu universaliste du droit naturel. Cf. Canto-Sperber [2005], p. 148-152.

4. Rousseau est ainsi obligé d'admettre, lors de ses réflexions sur les écrits de l'abbé de Saint-Pierre, que « d'homme à homme, nous vivons dans l'état civil et soumis aux loix ; de peuple à peuple, chacun jouit de la liberté naturelle : ce qui rend au fond notre situation pire que si ces distinctions étaient inconnues. » (EASP, p. 610.)



Existe-t-il cependant une échappatoire à la contrainte politique que nous venons d'évoquer ? La première possibilité, évoquée dans le *Discours sur l'économie politique*, consiste à envisager, à la suite de Diderot (Proust [1995], p. 388-390), la possibilité d'une volonté générale globale. La volonté d'un peuple unique constitué en tant que personne morale s'imposerait alors également à tous comme norme universelle de justice (*DEP*, p. 245). La possibilité d'un corps politique au niveau mondial, la *société générale*, est cependant difficilement envisageable dans le cadre théorique de la nation développé par Rousseau. L'idée d'une société générale est d'ailleurs fermement rejetée dans le *Manuscrit de Genève* (*MG*, p. 284). Une association globale des hommes n'étant ni possible ni souhaitable, une association de peuples, c'est-à-dire une *confédération*, permettrait selon lui d'associer la force du lien social qu'offre le patriotisme à l'universalité de la loi. Selon Rousseau, « excepté le Turc, il regne entre tous les Peuples de l'Europe, une liaison sociale imparfaite, mais plus étroite que les nœuds généraux et lâches de l'humanité » (*EASP*, p. 573). Il est donc possible d'envisager une « société libre et volontaire, qui unit tous les États Européens, prenant la force et la solidité d'un vrai Corps politique, [qui] peut se changer en une confédération réelle » (*EASP*, p. 574). Le premier enjeu de l'économie politique rousseauiste consiste donc à dépasser, par une confédération européenne, un cadre national qui se réduit invariablement à un retour à l'état de nature des corps politiques eux-mêmes. Mais, à l'instar des individus, les corps politiques ne peuvent s'inscrire dans une démarche politique fédératrice lorsqu'ils rivalisent d'orgueil et de vanité. Si à l'intérieur de la nation, l'hypothèse d'une rationalité économique « dominée » peut paraître acceptable<sup>1</sup>, en revanche elle ne peut être étendue au-delà du cadre de la nation. La notion de rationalité, qui trouve son fondement dans l'état de nature, a cependant d'abord une dimension universelle.

## Rationalité économique et mythe de l'autarcie

L'idée de rationalité dominée, avancée par Jean Mathiot, semble buter sur deux difficultés. La rationalité économique au sens strict, c'est-à-dire fondée sur une conception autonome de l'intérêt (*amour de soi*), est tout d'abord politiquement déterminante. Les individus, et à un niveau supérieur les nations, ont besoin de développer une rationalité économique substantielle, uniquement orientée vers la satisfaction des besoins, pour penser leur intérêt commun et envisager un accord politique. Le contrat social, ensuite, ne suffit pas nécessairement pour rétablir une rationalité de l'échange. Le poids historique des représentations collectives peut constituer un frein important à cette évolution. La représentation de l'autonomie économique des entités peut alors apparaître comme nécessaire à la normalisation de cette rationalité. Elle constitue *virtuellement* un moyen de transcender les représentations hétéronomes de l'intérêt. L'idée que nous voulons par conséquent défendre ici, c'est que la *représentation* de l'individu ou de la communauté (famille ou nation) comme entité moralement et économiquement autonome participe à la construction conceptuelle de la rationalité. L'état de nature et l'économie domestique représentent, à ce titre, des fictions identifiantes fondamentales.

1. Le terme de rationalité dominée est de Jean Mathiot ([2000], p. 28).

L'état de nature, ou plus exactement le discours rousseauiste sur l'état de nature, permet, tout d'abord, au lecteur de s'extraire virtuellement de son environnement pour se penser, et repenser les déterminants de son intérêt, en dehors de tout contexte social. En effet, « le plus sûr moyen de s'élever au-dessus des préjugés et d'ordonner ses jugemens sur les vrais rapports des choses est de se mettre à la place d'un homme isolé, et de juger de tout comme cet homme en doit juger lui-même eu égard à sa propre utilité » (Rousseau [1969], p. 291). L'introspection, que vise en fait la référence à l'état de nature, se propose de découvrir derrière le particularisme et la contingence d'un individu la véritable nature de l'homme. Rousseau, arguant de sa retraite de la civilisation, s'offre lui-même en modèle dans ses *Confessions* : « Je veux montrer à mes semblables un homme dans toute la vérité de la nature ; et cet homme ce sera moi. » (Rousseau [1968], p. 43.) L'éducation virtuelle d'Émile, en réalité du lecteur, de sa naissance à son mariage (métaphoriquement de l'état de nature à l'état civil) suit une logique similaire. La référence récurrente à Robinson Crusoé, seul livre recommandé pour l'éducation d'Émile, offre une grille de lecture à l'ouvrage éponyme et en constitue une ingénieuse mise en abîme<sup>1</sup>. Émile, personnage imaginaire et encore vierge de toute relation sociale, doit apprendre à s'identifier à Robinson pour s'assurer de l'autonomie réelle de son intérêt : « Robinson Crusôé dans son isle, seul, dépourvu de l'assistance de ses semblables et des instrumens de tous les arts, pourvoyant cependant à sa subsistance, à sa conservation, et se procurant même une sorte de bien-être, voilà un objet intéressant pour tout âge et qu'on a mille moyens de rendre agréable aux enfans. Voilà comment nous réalisons l'isle déserte qui me servoit d'abord de comparaison. Cet état n'est pas, j'en conviens, celui de l'homme social ; vraisemblablement il ne doit pas être celui d'Émile ; mais c'est sur ce même état qu'il doit apprécier tous les autres. » (Rousseau [1969], p. 291. C'est nous qui soulignons.)

L'économie domestique de Clarens, ensuite, longuement décrite dans *Julie, ou la Nouvelle Héloïse*, idéalise le retour à la vie rustique et familiale et symbolise ainsi le rejet de la superficialité du mode de vie bourgeois et de ses normes sociales (Starobinski [1971], p. 102-148 ; Philonenko [1984], t. II, p. 161-212 ; Markovits [1991]). Le contexte de *La Nouvelle Héloïse*, tout comme celui d'*Émile*, est la France monarchique, symbole de la civilisation policée mais corrompue dénoncée par Rousseau. Les personnages mis en scène par Rousseau tentent d'échapper à l'emprise morale de leur société. Clarens veut symboliser « un ordre de choses où rien n'est donné à l'opinion, où tout a son utilité réelle et qui se borne aux vrais besoins de la nature » (Rousseau [1964a], p. 547). La première caractéristique de l'économie de Clarens est donc l'autarcie. L'homme recrée ainsi un lien privilégié avec la nature, et c'est une condition essentielle de son autonomie morale (Starobinski [1971], p. 135 ; Taylor [2003], p. 297). Clarens est, en effet, une société quasiment fermée, autosuffisante et non monétaire, dans laquelle l'utilité réelle remplace l'usage et la convenance. L'élégance des parures de Julie et le raffinement des mets servis dans le salon d'Apolon rappellent au lecteur que l'autarcie n'est pas l'austérité, et que le bon goût n'est pas l'affaire de l'opinion (Rousseau [1964a], p. 549-550). Ils lui rappellent

1. On notera l'ironie de la référence. Robinson, sur son île, essaie, lui, de reproduire ou réinventer un monde « civilisé ». Notons également le *Télémaque* de Fénelon joue un rôle symétrique pour Sophie. Celle-ci tombe amoureuse de Télémaque (Émile) en s'identifiant à Eucharis (Rousseau [1969], p. 598).



également que le vrai luxe est dans la modération et la tempérance des passions<sup>1</sup>. Clarens constitue, à ce titre, une forme renouvelée et élargie de robinsonnade.

L'univers fictionnel, voire utopique, de Rousseau se structure logiquement autour de l'autarcie. L'objet du discours rousseauiste, et plus généralement des sciences sociales, consiste à donner aux hommes la connaissance de soi, cette faculté de jugement et d'imagination, dont les hommes ont besoin pour transcender les liens, nécessaires mais trop étroits, de la famille et de la nation (Levi-Strauss [1962], p. 243). L'autarcie virtuelle de l'économie domestique ne se présente donc pas comme une idéalisation de l'autarcie réelle de l'économie politique, elle tente, au contraire, de la transcender. Elle appartient, au même titre que l'état de nature, à un ensemble théorique (et rhétorique) destiné à briser le cercle dit de l'autoposition. La difficulté logique à laquelle Rousseau fait en effet face dans le *Contrat social*, à savoir que les hommes doivent être avant la loi ce qu'ils doivent devenir par elle<sup>2</sup>, représente la limite du politique dans la réforme morale des hommes. Le discours peut y suppléer en incitant les hommes à la vertu. La forme du discours, l'usage de la fiction notamment, est à ce titre essentielle, et on aurait sans doute tort de la négliger au profit des seuls essais politiques<sup>3</sup>. À l'origine simple médiation, le langage, à l'instar de l'échange, a été perverti en instrument de domination. Rousseau est donc lui-même contraint par le langage conventionnel (Starobinski [1971], p. 179-181) qui ne lui permet d'exprimer qu'imparfaitement la vraie nature de l'homme. La fiction permet, grâce à l'imagination et l'émotion qu'elle provoque, d'y pallier partiellement. En effet, plus l'identification est forte, moins la médiation du discours est destructrice de sens. La figure et le style sont donc déterminants dans le processus d'identification<sup>4</sup>. Rousseau témoigne d'ailleurs du rôle initiatique qu'ont joué les romans dans sa jeunesse (Rousseau [1968], p. 46). Il reconnaît donc à l'identification littéraire (« je veux [...] qu'il pense être Robinson lui-même ») un rôle indispensable à toute réforme anthropologique et politique. Selon Martin Rueff, « promue méthode de recherche, la fiction est le lieu même du système philosophique : de même que dans le *Second Discours*, elle suggérait des routes oubliées de l'état de nature, elle collabore ici à l'engendrement de l'homme nouveau » (Rueff [2000], p. 298). Nous ne partageons donc pas la thèse de Robert Derathé pour qui l'*Émile* et la *Nouvelle Héloïse* d'une part, le *Contrat social*, le *Projet de constitution pour la Corse* ou les *Considérations sur le gouvernement de*

1. Julie aime-t-elle le café ? Elle en perd l'habitude « pour en augmenter le goût [...] C'est une petite sensualité qui la flate plus, qui lui coûte moins, et par laquelle elle aiguise et règle sa gourmandise » (Rousseau [1964a], p. 552). Mais, bien que le luxe soit officiellement banni à Clarens, il y est imité. Les grands crus de vin sont ainsi remplacés par le vin produit sur place, mais servi sous de fausses étiquettes (Rousseau [1964a], p. 552-53). Sur la différence entre volupté et vanité, voir aussi Rousseau ([1969], p. 514-527).

2. « Pour qu'un peuple naissant put goûter les saines maximes de la politique et suivre les règles fondamentales de la raison d'État, il faudrait que l'effet put devenir la cause, que l'esprit social qui doit être l'ouvrage de l'institution présidât à l'institution même, et que les hommes fussent avant les Loix ce qu'ils doivent devenir par elles. » (CS, II, p. 383.)

3. Le succès éditorial de *La Nouvelle Héloïse* ou de l'*Émile* fut d'ailleurs énorme en comparaison du *Contrat social*, Belin [1913], p. 160.

4. « La spontanéité de l'écriture, calquée en principe sur la spontanéité du sentiment actuel (lequel se donne comme une émotion ancienne revécue) assure l'absolue authenticité de la narration. Le style, au dire même de Rousseau, prend dès lors une importance qui ne se limite plus à la seule mise en œuvre du langage, à la seule recherche technique des effets : il devient emphatique "auto-référentiel", il prétend renvoyer immanquablement à la vérité "intérieure" de l'auteur. » Starobinski [2001], p. 122.

*Pologne* de l'autre, appartiendraient à deux démarches et s'adresseraient à deux types de pays différents. Le genre littéraire s'adresserait aux habitants des grands États inaptes à l'institution légitime et le genre politique aux États susceptibles de se réformer ou de corriger leur constitution. Les premiers y rechercheraient les voies du bonheur, les seconds les préceptes de la vertu (Derathé [1962], p. 218). En dépit de l'hétérogénéité de leurs genres discursifs, nous croyons au contraire que tous ces écrits se rejoignent sur un point. Ils tendent tous à diffuser une nouvelle représentation de l'intérêt autonome propre à réformer les institutions. Ils offrent une norme de la rationalité centrée sur une détermination réellement autonome des individus et une maîtrise parfaite de leurs passions sociales telles que l'envie, l'honneur ou l'orgueil. La vertu civile, c'est-à-dire l'adéquation des intérêts particuliers à l'intérêt général (*DEP*, p. 252), et le bonheur ne sont possibles qu'à cette condition.

## CONCLUSION

La question de la rationalité, que nous avons tenté de présenter ici au travers des rapports entre intérêt et loi, mériterait certainement des développements beaucoup plus importants. Nous avons choisi de la traiter autour du thème de l'échange que Rousseau aborde presque systématiquement de manière négative. Il ressort néanmoins de cette entreprise que la rationalité économique qui le sous-tend fait l'objet de toute son attention. L'échange n'est selon lui responsable de la dépendance sociale, et des maux qui en résultent, que lorsqu'il répond à une rationalité autre qu'économique. C'est le cas, en particulier, lorsque l'intérêt qui motive la transaction n'est pas déterminé de manière autonome. L'incapacité des individus à se comporter en fonction de leurs véritables besoins est également à l'origine de la perturbation des mécanismes économiques eux-mêmes. Le secteur agricole, qui devrait être prioritaire pour Rousseau, est alors défavorisé au profit du commerce ou de la finance. L'économie justifie donc à tort des pratiques dont on peut légitimement questionner la rationalité. C'est paradoxalement la loi qui assure théoriquement la rationalité de l'échange. Deux difficultés se présentent cependant, le poids historique des représentations collectives et la limite politique de la nation. En dépit de l'égalité civile que leur donne la loi, les citoyens peuvent tout d'abord continuer à voir dans la richesse un instrument de pouvoir et d'honneur. L'échange entre nations, ensuite, ne peut se prévaloir de l'influence de la loi. L'autarcie constitue alors légitimement une mesure de prudence. Dans l'un et l'autre cas, l'autonomie de l'intérêt doit cependant pouvoir être sauvegardée (ou restaurée) en complément de (ou autrement que par) la loi. C'est précisément le rôle du discours, et notamment de la fiction, plus apte à modifier les représentations collectives d'une société que l'essai. Les fictions identifiantes de *l'Émile* ou de *Julie*, auxquelles on peut ajouter les écrits autobiographiques de Rousseau, permettent ainsi aux lecteurs de se recentrer sur leurs besoins et leur plaisir indépendamment de l'opinion. L'autarcie virtuelle peut se lire comme une expérience de pensée destinée, non pas à restreindre l'échange, mais au contraire à en fonder la légitimité. Elle illustre alors très bien l'idée développée par Jean Starobinski, et à laquelle il donna le titre de l'un de ses ouvrages, *Le Remède dans le mal*.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- AUBENQUE P., [1963], *La prudence chez Aristote*, Paris, PUF.
- BACH R. (dir.), [1999], *Rousseau et l'économie politique, Études J.-J. Rousseau*, n° 10.
- BACHOFEN B., [2002], *La condition de la liberté. Rousseau, critique des raisons politiques*, Paris, Payot.
- BAUD-BOVY S., DERATHÉ R., DOTRENS R. et al., [1962], *Jean-Jacques Rousseau*, Neuchâtel, À La Baconnière.
- BELIN J.-P., [1913], *Le mouvement philosophique de 1748 à 1789*, Paris, Belin Frères.
- BERNARDI B. (dir.), [2002], *Jean-Jacques Rousseau. Discours sur l'économie politique*, Paris, Vrin.
- BOULAD-AYOUB J., SCHULTE-JENCKHOFF I. et VERNES P.-M. (dir.), [2000], *Rousseau. Anticipateur-retardataire*, Paris, L'Harmattan.
- BOUVIGNIES I., [2002], « Droit de propriété et domaine public », dans BERNARDI B. (dir.), *Jean-Jacques Rousseau. Discours sur l'économie politique*, Paris, Vrin, p. 173-193.
- CHARLES L., STEINER P., [1999], « Entre Montesquieu et Rousseau. La physiocratie parmi les origines intellectuelles de la Révolution française », dans *Études Jean-Jacques Rousseau*, n° 10, p. 83-160.
- CANTO-SPERBER M., [2005], *Le bien, la guerre & la terreur. Pour une morale internationale*, Paris, Plon.
- DERATHÉ R., [1950], *Jean-Jacques Rousseau et la science politique de son temps*, Paris, Vrin, rééd. 1995.
- DERATHÉ R., [1962], « L'unité de la pensée de Jean-Jacques Rousseau », dans BAUD-BOVY S., DERATHÉ R., DOTRENS R. et al., *Jean-Jacques Rousseau*, Neuchâtel, À La Baconnière, p. 203-218.
- EYSSIDIEUX-VAISSERMANN A., [2001], « Rousseau et la science de l'économie politique dans l'*Encyclopédie* », *Kairos*, 18, p. 47-73.
- FRIDÉN B., [1998], *Rousseau's economic philosophy. Beyond the Market of Innocents*, Dordrecht, Kluwer Academic Publishers.
- GOLDSCHMIDT V., [1974], *Anthropologie et politique. Les principes du système de Rousseau*, Paris, Vrin, rééd. 1983.
- LARRÈRE C., [1992], *L'invention de l'économie au XVIII<sup>e</sup> siècle*, Paris, PUF.
- LEVI-STRAUSS C., [1962], « J.-J. Rousseau, fondateur des sciences de l'homme », dans BAUD-BOVY S., DERATHÉ R., DOTRENS R. et al., *Jean-Jacques Rousseau*, Neuchâtel, À La Baconnière, p. 239-248.
- MARKOVITS F., [1986], *L'ordre des échanges. Philosophie de l'économie et économie du discours au XVIII<sup>e</sup> siècle en France*, Paris, PUF.
- MARKOVITS F., [1991], « Rousseau et l'éthique de Clarens : une économie des relations humaines », *Stanford French Review*, p. 323-348.
- MASTER R.D., [1968], *The Political Philosophy of Rousseau*, Princeton et Lyon, ENS Éditions, 2002 pour la traduction française.
- MATHIOT J., [2000], « Politique et économie chez Jean-Jacques Rousseau », dans J. BOULAD-AYOUB, I. SCHULTE-JENCKHOFF et P.-M. VERNES (dir.), *Rousseau. Anticipateur-retardataire*, Paris, L'Harmattan, p. 19-39.
- PETTIT P., [1997], *Republicanism. A Theory of Freedom and Government*, Oxford University Press, et Paris, Gallimard, 2004 pour la traduction française.
- PHILONENKO A., [1984], *Jean-Jacques Rousseau et la pensée du malheur*, 3 tomes, Paris, Vrin.
- ROUSSEAU J.J., [1969], *Émile ou De l'éducation*, Paris, Gallimard, rééd. 2002 (Folio Essais).
- ROUSSEAU J.J., [1968], *Les Confessions*, Paris, Garnier Flammarion, 2 volumes, rééd. 2002, GF Flammarion.
- ROUSSEAU J.J., [1967], *Lettre à M. d'Alembert sur son article Genève*, Paris, Garnier Flammarion.

- ROUSSEAU J.-J., [1964a], *Œuvres complètes*, t. II, *La Nouvelle Héloïse. Théâtre. Essais littéraires*, Paris, Gallimard.
- ROUSSEAU J.-J., [1964b], *Œuvres complètes*, t. III, *Du Contrat social. Écrits politiques*, Paris, Gallimard.
- RUEFF M., [2000], « L'ordre et le système : l'empirisme réfléchissant de Jean-Jacques Rousseau », dans J. BOULAD-AYOUB, I. SCHULTE-JENCKHOFF et P.-M. VERNES (dir.), *Rousseau. Anticipateur-retardataire*, Paris, L'Harmattan, p. 275-344.
- SALVAT C., [2006], « *Economies*. Les articles "Œ/Economie" et leurs désignants dans l'*Encyclopédie* », in *Recherches sur Diderot et l'Encyclopédie*, 40, p. 97-116.
- STEINER P., [1998], *La « science nouvelle » de l'économie politique*, Paris, PUF.
- SCUBLA L., [1992], « Est-il possible de mettre la loi au dessus de l'homme ? Sur la philosophie politique de Jean-Jacques Rousseau », dans J.-P. DUPUY, *Introduction aux sciences sociales. Logique des phénomènes collectifs*, Paris, Ellipses, p. 105-143.
- SKINNER Q., [1998], *Liberty before liberalism*, Cambridge University Press/*La liberté avant le libéralisme*, Paris, Seuil, 2000.
- SPECTOR C., [2003], « Rousseau et la critique de l'économie politique », dans B. BENSUADE-VINCENT et B. BERNARDI (dir.), *Rousseau et les sciences*, Paris, L'Harmattan.
- SPITZ J.-F., [1995], *La liberté politique. Essai de généalogie conceptuelle*, Paris, PUF.
- STAROBINSKI J., [1971], *Jean-Jacques Rousseau. La transparence et l'obstacle*, Paris, Gallimard, rééd. 2000.
- STAROBINSKI J., [2001], *La relation critique*, Paris, Gallimard.
- STAROBINSKI J., [1999], *L'œil vivant*, Paris, Gallimard.
- STAROBINSKI J., [1989], *Le remède dans le mal. Critique et légitimation de l'artifice à l'âge des Lumières*, Paris, Gallimard.
- VIROLI M., [1988], *Jean-Jacques Rousseau and the "well-ordered society"*, Cambridge, Cambridge University Press.





## Recherches sur Diderot et sur l'Encyclopédie

40-41 | octobre 2006

Les branches du savoir dans l'Encyclopédie

---

### Les articles 'Économie' et leurs désignants

Christophe Salvat

---



#### Édition électronique

URL : <http://journals.openedition.org/rde/352>

DOI : 10.4000/rde.352

ISSN : 1955-2416

#### Éditeur

Société Diderot

#### Édition imprimée

Date de publication : 1 octobre 2006

Pagination : 107-126 Nous tenons à remercier les organisateurs et les intervenants du colloque sur les désignants, ainsi que M.F. Piguet, P. Crépel et C. Théré pour leurs commentaires et leurs critiques de cet article.

ISBN : 2-952089-6-4

ISSN : 0769-0886

#### Référence électronique

Christophe Salvat, « Les articles 'Économie' et leurs désignants », *Recherches sur Diderot et sur l'Encyclopédie* [En ligne], 40-41 | octobre 2006, mis en ligne le 01 octobre 2008, consulté le 20 avril 2019. URL : <http://journals.openedition.org/rde/352> ; DOI : 10.4000/rde.352

---

Ce document a été généré automatiquement le 20 avril 2019.

Propriété intellectuelle

---

# Les articles 'Œ\Économie' et leurs désignants

Christophe Salvat

---

- 1 Le regard de l'historien de la pensée économique sur l'*Encyclopédie* est, le plus souvent, attiré par les deux contributions séminales, que sont les articles FERMIERS et GRAINS de Quesnay, ce dernier étant considéré comme l'un des principaux fondateurs de la discipline. Sa curiosité est également attisée par un événement linguistique majeur, l'usage jusque-là rarissime de la dénomination 'moderne' d'économie politique, en tant que désignant de ces deux articles. L'affaire semble claire, l'économie politique, en tant que discipline nouvelle, serait née dans l'*Encyclopédie*.
- 2 Marie-France Piguet, dans un article publié récemment<sup>1</sup>, a cependant montré toute la complexité inhérente aux dénominations économiques dans l'*Encyclopédie*. Se concentrer sur le seul usage de *économie politique*, *a fortiori* dans deux articles seulement, est une démarche trop réductrice. S'interroger sur l'originalité de l'emploi du syntagme *économie politique*, et sur sa signification, exige d'interroger préalablement *œ/économie* et ses composés : économie rustique, animale, domestique, politique... Marie-France Piguet a ainsi pu mettre en évidence la rareté relative du syntagme 'économie politique' et de la grande polysémie du terme *œ/économie*. Notre étude s'inscrit directement dans la continuité de son article, et se propose plus spécifiquement d'identifier deux sémantiques de l'économie politique, articulées chacune autour d'un syntagme lexicalement identifié, l'économie domestique et l'économie rustique.
- 3 Nous nous proposons donc ici de revenir, dans un premier temps, sur la tradition lexicographique du terme *œ/économie* afin de mieux identifier la nouveauté de l'*Encyclopédie* en la matière, puis sur les emplois propres à cette dernière, les articles et les désignants *œ/économie*. Dans une seconde partie, nous envisagerons l'*économie publique* et/ou *politique* dans ses rapports à l'*économie domestique* et l'*économie rustique*. Nous montrerons ainsi que Rousseau défend une économie entendue comme mode d'administration d'une communauté, qu'elle soit familiale ou nationale, alors que Quesnay et Leroy la perçoivent davantage comme mode de production (agricole), au

niveau individuel comme au niveau national. Bien qu'appartenant à des perspectives différentes, ces deux approches peuvent fortuitement se recouper et/ou s'opposer, par exemple sur la question des flux monétaires, de l'industrie agricole ou de la fiscalité, mais elles ne peuvent être considérées comme des théories économiques concurrentes.

## 1. Lexicographie et encyclopédisme

### a. Les sources de l'*Encyclopédie*

- 4 L'édition de 1690 du *Dictionnaire universel* de Furetière définit le substantif 'œconomie' comme suit :
 

ŒCONOMIE. Subst. Masc. [sic ?] Mesnagement prudent qu'on fait de son bien, ou de celui d'autrui. L'*œconomie* est la seconde partie de la Morale, qui enseigne à bien gouverner une famille, une Communauté. Ce prier entend bien l'*œconomie*. Quelquefois on couvre l'avarice du nom honneste d'*œconomie*.

ŒCONOMIE, signifie quelquefois, bel ordre & disposition des choses. C'est une chose admirable que l'*œconomie* & la disposition des parties du corps humain, & de voir comme chacune fait ses fonctions. On dit aussi l'*œconomie* d'un bastiment, en parlant du mesnagement de la place, & de la belle & commode disposition des appartements. On admire aussi l'*œconomie* des abeilles, & la belle maniere dont elles se gouvernent dans leur petite Republique<sup>2</sup>.
- 5 Sa forme adjectivée, *œconomique*, renvoie explicitement aux écrits apocryphes d'Aristote<sup>3</sup>. Cette définition servira de base aux entrées postérieures du dictionnaire de Trévoux et de l'*Encyclopédie*. On y retrouve les trois principaux sens du terme, à savoir l'administration (familiale) des biens, l'ordre d'un ensemble et le gouvernement d'une communauté. L'*oe/économie* désigne donc communément et indistinctement la conduite d'un ménage ou d'une société politique.
- 6 Le dictionnaire de Trévoux fut, ainsi que l'a montré Marie Leca-Tsiomis<sup>4</sup>, une source fondamentale pour l'*Encyclopédie*. Il développe et organise les différentes acceptions du terme supportées par des citations. L'article ÉCONOMIE comprend une définition générale et deux sous-entrées ÉCONOMIE. À l'intérieur de celles-ci sont insérées des variantes ou des locutions, qui sont, surtout dans l'édition de 1771, indiquées par un petit pictogramme en forme de main. L'édition de 1752 définit, dans sa forme générale, l'économie comme la « conduite sage, [le] ménagement prudent qu'on fait de son bien ou de celui d'autrui.[...] L'économie est la seconde partie de la Morale, qui enseigne à bien gouverner une famille, une communauté »<sup>5</sup>. L'exemple type d'une bonne économie, celle du père de famille, est donné en référence à Bossuet. L'édition de 1771 sera plus explicite sur ce que doit être une bonne gestion familiale<sup>6</sup>. La locution ÉCONOMIE RUSTIQUE y sera également rattachée.
- 7 La première sous-entrée 'ÉCONOMIE' présente ensuite une première variation sémantique. Par extension, l'économie désigne également le « bon usage qu'on fait de son esprit, & de ses autres qualités : de la prudence à les bien placer, ou à les bien ménager. *Œconomia, prudens administratio* »<sup>7</sup>. Suivent ensuite les locutions ÉCONOMIE LÉGALE<sup>8</sup> et ÉCONOMIE ÉVANGÉLIQUE<sup>9</sup>. Le chevalier de Jaucourt reprendra à son compte, en les synthétisant, ces définitions dans ÉCONOMIE (*Critiq. sacrée.*)<sup>10</sup>.
- 8 La sixième édition comprend également une seconde extension : l'économie « dans une signification plus étendue, se dit, au figuré, de l'ordre par lequel un corps politique subsiste principalement. Renverser toute l'*économie* d'un État »<sup>11</sup>. Cette acception est



absente dans l'édition de 1732. Dans l'édition de 1718 du *Dictionnaire de l'Académie française* (les termes d'*économie* ou d'*œconomie* sont absents de l'édition de 1694), l'*œconomie* désigne cependant déjà l'organisation du corps politique : « L'ordre, la règle qu'on apporte dans la conduite d'un ménage, dans la dépense d'une maison. [...] se dit figurément de l'ordre par lequel un corps politique subsiste principalement »<sup>12</sup>.

- 9 La seconde sous-entrée de l'article est essentiellement un dérivé métonymique de la précédente :

ÉCONOMIE, signifie quelquefois le bel ordre, & la juste disposition des choses. *Concinna series*. C'est une chose admirable que l'*économie* & la disposition des parties du corps humain, & de voir comme chacune fait régulièrement ses fonctions. On dit aussi l'*économie* d'un bâtiment, en parlant du ménagement de la place, & de la belle & commode disposition des appartemens. L'*économie* d'un tableau. *Économie* du dessein. Une belle *économie*. On admire aussi l'*économie* des abeilles, & la belle manière dont elles se gouvernent dans leur petite république. On dit aussi, par la même raison, l'*économie* d'un État.<sup>13</sup>

- 10 On y retrouve partiellement la rédaction de Furetière, notamment l'exemple des abeilles, ici étendu à l'État. Dans un premier temps, en 1752, Trévoux considère l'*économie* d'un corps politique en tant qu'ordre ou harmonie d'un ensemble, *Concinna series*, alors qu'en 1771, il l'associe à la *prudens administratio*. Nous pouvons légitimement y voir l'influence de l'article ÉCONOMIE de Rousseau. La notion d'*économie animale*, ajoutée en 1771, appartient sans ambiguïté à l'*économie Concinna series*.
- 11 Si les 4<sup>e</sup> et 5<sup>e</sup> éditions (respectivement 1743 et 1752) du dictionnaire de Trévoux ont sans aucun doute été exploitées pour la rédaction des articles de l'*Encyclopédie*, il est donc également juste de noter que la 6<sup>e</sup> édition du dictionnaire, publiée en 1771, s'est aussi beaucoup inspiré des avancées de l'*Encyclopédie*. On le note en tout cas pour l'article ÉCONOMIE. Les deux nouvelles locutions qui font leur apparition dans la 6<sup>e</sup> édition, l'*économie rustique*<sup>14</sup> et l'*économie animale*<sup>15</sup> semblent, en effet, beaucoup devoir leur rédaction à celle de l'*Encyclopédie*, quoique l'hypothèse d'une troisième source ne soit théoriquement pas à exclure<sup>16</sup>.

## b. Les articles 'ECONOMIE' et 'ŒCONOMIE' dans l'*Encyclopédie*

- 12 L'*économie* fait l'objet de deux entrées dans l'*Encyclopédie*, usant ainsi de la polygraphie du terme, pour compléter ou corriger dans le tome XI ce qui avait été publié, dix ans plus tôt, dans le tome V. La première entrée se compose du célèbre article ÉCONOMIE (*Morale et Politique*) de Rousseau, réédité en 1758 sous le titre *Discours sur l'économie politique*, et d'une sous-entrée ÉCONOMIE RUSTIQUE.
- 13 La seconde entrée du mot, à la lettre O, répond davantage au souhait exprimé par Diderot dans l'article ENCYCLOPÉDIE de mieux intégrer les termes de la langue usuelle<sup>17</sup>. Elle se décline en quatre parties, la définition générale, et trois sous-entrées, ŒCONOMIE (*Critiq. sacrée*) de Jaucourt, ŒCONOMIE ANIMALE (*Médec.*) de Ménuret de Chambaud, et ŒCONOMIE POLITIQUE (*Hist. Pol. Rel. ant. & mod.*) de 'feu Boulanger'. La définition générale renvoie aux entrées du tome V : « Œconomie, conduite sage et prudente que tient une personne en gouvernant son propre bien ou celui d'un autre. Il y a l'*œconomie politique*. Voyez ce mot à l'orthographe Économie. Il y a l'*œconomie rustique*; c'est ce qui a rapport à toute la vie rustique »<sup>18</sup>. En réintégrant les deux sous-entrées du tome V, l'article ŒCONOMIE répond donc parfaitement à l'enjeu lexicographique de l'*Encyclopédie*.

- 14 L'article de Rousseau 'ÉCONOMIE ou ŒCONOMIE (*Morale & Politique*)' a souvent été abusivement désigné par ÉCONOMIE POLITIQUE<sup>19</sup>, et ce à l'intérieur même de l'*Encyclopédie*<sup>20</sup>. En réalité, Rousseau n'emploie jamais l'expression d'*économie politique* dans cet article, mais celui d'*économie publique*, qu'il utilise à sept reprises. Une hésitation, cependant, au début de son article : « économie générale, ou politique »<sup>21</sup>, à laquelle Rousseau ne donnera pas suite. On doit en fait à Jacob Vernes, et à son édition de 1758 du *Discours sur l'économie politique*, la confusion entre l'article ÉCONOMIE et l'expression d'économie politique. Rousseau ne semble pas cependant avoir été associé au choix du titre<sup>22</sup>. L'attribution abusive d'un article ÉCONOMIE POLITIQUE à Rousseau, par Diderot lui-même, n'est probablement pas étrangère à l'intitulé de l'article ŒCONOMIE POLITIQUE composé à partir des papiers de Boulanger.
- 15 L'article ŒCONOMIE POLITIQUE fut publié en 1765, soit six ans après la mort de son signataire posthume, Nicolas-Antoine Boulanger. Il n'y a cependant aucun doute sur son attribution. L'article est une refonte, composée de larges extraits, du manuscrit des *Recherches sur l'origine du despotisme oriental* qui sera également publié à titre posthume en 1761<sup>23</sup>. L'intitulé n'est cependant pas de Boulanger qui n'utilise l'expression *oeconomie politique* à aucun moment dans le *Despotisme oriental*, pas plus que ne l'est la définition qui lui est accolée :
- « ŒCONOMIE POLITIQUE (*Hist. Pol. Rel. anc. & mod.*) c'est l'art & la science de maintenir les hommes en société, & de les y rendre heureux, objet sublime, le plus utile & le plus intéressant qu'il y ait pour le genre humain » (ŒCONOMIE POLITIQUE, XI, 366-67).
- 16 Cette définition fait écho à l'article SOCIÉTÉ dans lequel le principe général, ou l'*économie*, de toute société, est défini par le bonheur des hommes<sup>24</sup>. Notons par ailleurs que le texte de l'*Encyclopédie* a substitué l'expression *œconomie publique* à celle d'*œconomie civile* employée par Boulanger dans le *Despotisme oriental*<sup>25</sup>.
- 17 L'*œconomie politique* s'entend ici comme la règle, la constitution, des sociétés humaines. De fait, Boulanger traite des différents types de gouvernement, la république, le despotisme, la monarchie auxquels il ajoute la théocratie qu'il considère à la fois comme le fondement et la source de tous les maux des régimes modernes. Bien qu'adoptant une démarche tout à fait opposée à celle de Rousseau<sup>26</sup>, cet article complète ÉCONOMIE plus qu'il ne le critique. Le manuscrit du *Despotisme oriental* fut d'ailleurs très probablement rédigé à l'automne 1755<sup>27</sup>. La véritable critique de l'article de Rousseau doit être recherchée dans l'article VINGTIÈME<sup>28</sup>.

### c. Œ /Économie rustique, domestique, animale et politique : les désignants économiques

- 18 L'étude menée par Marie-France Piguet<sup>29</sup> lors d'un précédent colloque sur l'*Encyclopédie* montre que sur les 1004 occurrences répertoriées du terme *œ/économie*, 9 seulement concernent l'*œ/économie politique*, loin derrière l'*œ/économie animale* qui représente 40 % des emplois (406). On doit rajouter à cela les formes tronquées du terme, 27 pour *œ/économie politique*, mais qui ne modifient pas la hiérarchie affichée pour les formes pleines. Les formes tronquées, du type *écon. polit.* ou *économie polit.*, sont les formes usuelles des désignants. On compte également dans l'*Encyclopédie* 9 occurrences de *économie générale*, 6 de *économie nationale*, 13 de *économie publique* dont 7 chez Rousseau<sup>30</sup>.

- 19 L'expression *œ/économie politique* est donc très marginalement employée sur les 74000 articles que compte l'*Encyclopédie*. Mais, et il s'agit là de l'un des principaux résultats de cette étude, elle est essentiellement utilisée pour désigner un nom de domaine puisqu'on compte 10 entrées ou sous-entrées rapportées au domaine *œ/économie politique* (contre 219 pour *œ/économie rustique*, 33 pour *œ/économie domestique*, 27 pour *œ/économie animale* et 10 pour *œ/économie*). Cette caractéristique est également valable pour *œ/économie domestique*. Les expressions *œ/économie domestique* ou *politique* sont donc peu usitées dans le corps des articles, et davantage dans l'appareil encyclopédique des désignants et des renvois. L'usage et la diffusion du syntagme *économie politique* doivent donc être interrogés dans la perspective globale du projet éditorial de l'*Encyclopédie*.
- 20 Dix articles ont pour désignant *économie politique*. Il s'agit de FINANCES de Pesselier, INTÉRÊT, MENDIANT, RENTIER, SEL de Jaucourt, FERMIERS et GRAINS de Quesnay et VINGTIÈME signé Boulanger mais en partie de Diderot<sup>31</sup>. À ceux-ci on peut rajouter deux petits articles à désignants composés, RASPHUIS<sup>32</sup> (*Hist. mod. Economic politiq.*) et SPINHUYS<sup>33</sup> (*Hist. mod. Econom. politiq.*). FINANCES est lui-même désignant d'une cinquantaine d'articles. Les articles de Véron de Forbonnais ont pour la plupart tous été publiés dans les cinq premiers volumes de l'*Encyclopédie*, soit avant le premier emploi du désignant *économie politique* pour l'article FERMIERS de Quesnay. Ils sont néanmoins liés à différents niveaux aux articles 'économiques'. Les articles ÉCONOMIE RUSTIQUE et FERMIER (*économie rustique*), renvoient ainsi à CULTURE DES TERRES (*commerce politique*) qui renvoie lui-même à AGRICULTURE de Diderot et à GRAINS (*économie politique*) de Quesnay. L'article COMMERCE également de Forbonnais renvoie également à AGRICULTURE, qui renvoie lui-même à différents articles d'économie rustique (FEUILLE, FROMENT, SEMOIR) ainsi qu'à l'article GRAIN (sic).
- 21 On observe de la part des éditeurs de l'*Encyclopédie* une volonté de ne pas employer le seul terme d'*œconomie* (une dizaine d'occurrences sur trois cents) au profit des désignants composés *économie rustique* et *économie domestique*. L'article FORET de Leroy constitue l'une des rares exceptions<sup>34</sup>. Les désignants *économie rustique*, parfois abrégés par *économie* (ex. FARINE, GRANGE) renvoient généralement à un ensemble de techniques productives de l'agriculture, parfois fondées sur les dernières recherches scientifiques, et non au savoir familial et ménager de l'*économie domestique*.
- 22 Les désignants *domestique* sont peu nombreux, une trentaine, et marquent le plus souvent des articles ménagers (CORBEILLE, CRUCHE, ÉTEIGNOIR, GOBELET, MOUCHOIR...). L'économie domestique, dans l'emploi qui en est fait dans la désignation des articles, est – on le voit – assez lointaine de l'approche aristotélicienne de Rousseau. Les articles d'économie rustique, le plus souvent rédigés par Diderot, représentent les trois quarts des articles dits d'*économie* dans l'*Encyclopédie*. On en compte environ 220, dont AGNEAU, CHASSE, ENGRAIS, FERMIER, LABOUR, METAYER, SEMAILLE ou TROUPEAUX.
- 23 Il n'est évidemment pas possible, dans le cadre d'un simple article, d'étudier l'ensemble de ces articles. Une illusoire tentative de synthèse des articles ayant pour intitulé ou désignant des composés d'*œ/économie* ne parviendrait d'ailleurs pas à donner une représentation honnête de l'économie au milieu du dix-huitième siècle. Chacun sait, en effet, que la grande majorité des analyses – que l'on qualifierait aujourd'hui d'*économiques* – était alors pensée en terme de *commerce* et de *commerce politique*<sup>35</sup>. En revanche, nous nous proposons ici de questionner la polysémie du syntagme *économie politique*, au sein d'un article ou en tant que désignant, dans ses relations à l'*économie rustique* et *domestique*.

C'est au travers de ces articulations, croyons nous, que nous devons appréhender les contributions majeures de Quesnay et de Rousseau.

## 2. L'économie politique et ses doubles

### a. De l'économie rustique à l'économie politique : les deux articles FERMIER(S)

- 24 Il existe deux articles FERMIER(S), l'un signé de Leroy et désigné par *économie rustique* (vedette au singulier), l'autre signé de Quesnay et désigné par *économie politique* (vedette au pluriel). Ce dernier occupe une place considérable pour les historiens de la pensée économique français qui trouvent là le premier écrit économique de l'un des fondateurs de leur discipline et l'une des toutes premières occurrences de sa dénomination. Dans le cadre de notre étude, il présente également une spécificité essentielle : il s'agit de la première utilisation du syntagme en tant que désignant. Pourquoi a-t-il été choisi pour cet article là ? L'expression *économie politique* semble peu courante à l'époque. Elle n'apparaît, à notre connaissance, dans aucun dictionnaire français avant l'*Encyclopédie* de Diderot et D'Alembert. Son origine n'est pas d'ailleurs clairement déterminée<sup>36</sup>. Il n'est employé qu'une seule fois dans les cinq premiers volumes de l'*Encyclopédie*<sup>37</sup>. A-t-on voulu faire référence par ce choix à l'article de Rousseau ? Cela semble peu probable<sup>38</sup>.
- 25 Une seconde hypothèse, plus probante selon nous, consiste – dans le cas présent – à comprendre le désignant non comme une appartenance disciplinaire mais comme une distinction lexicographique. Autrement dit, *économie politique* aurait été choisi pour distinguer l'article FERMIERS de Quesnay de l'article précédent de Leroy logiquement désigné par *économie rustique*. Les deux articles sont complémentaires l'un de l'autre. Quesnay, qui ne cite généralement que ses propres articles, renvoie ainsi son lecteur à l'article de Leroy<sup>39</sup>.
- 26 Dans l'article FERMIERS, Quesnay oppose le fermage au métayage. Il développe une représentation dualiste de l'agriculture associant mode d'exploitation et mode de production ; les terres cultivées par des fermiers le sont avec des chevaux alors que celles qui sont cultivées avec des métayers le sont avec des boeufs. À l'instar de Leroy, Quesnay veut se distinguer d'un certain nombre de savoirs ou de croyances populaires, « des idées vagues et imparfaites » de ceux qui commentent l'état de l'agriculture « sous un aspect général » alors qu'on « ne peut connaître ces objets que par un examen fort étendu et fort approfondi ». Le texte est ainsi parsemé de périphrases marquant la distanciation ainsi établie entre le savoir approfondi, 'scientifique' de l'auteur, et les représentations vulgaires et erronées du peuple<sup>40</sup>. Par opposition, le discours de Quesnay se présente comme rigoureux, analytique et démonstratif<sup>41</sup>. D'un point de vue *politique*, la petite exploitation agricole familiale traditionnelle, celle du métayage, celle de la quasi-autarcie, est à proscrire. La petite culture se fait en effet au détriment de l'État.
- 27 L'*économie rustique*, comme l'*économie politique*, rompt avec les représentations communes ou traditionnelles du 'bon oe/économe', des principes d'oe/économie du bon père de famille, ou encore la *science économique* de Faiguet de Villeneuve<sup>42</sup> : « Dans la maison on ne gagne qu'en épargnant, dans le champ une grande hardiesse à dépenser est souvent nécessaire pour gagner beaucoup. Il arrive très-souvent que les *fermieres* qui deviennent veuves, se ruinent, parce qu'elles conduisent toute la ferme par les principes qui ne

conviennent qu'à la basse-cour »<sup>43</sup>. La ferme, écrit encore Leroy à l'article éponyme, ne doit pas être confondue avec la 'maison rustique'<sup>44</sup>.

- 28 L'économie rustique se présente explicitement dans les écrits de Leroy comme un savoir agronomique destiné à améliorer le rendement agricole. Celui-ci s'appuie parfois sur les autorités de scientifiques tels que Buffon ou Réaumur, par exemple à l'occasion du débat sur les baliveaux dans l'article FORET. Mais les fermiers, loin d'être d'ignares paysans, sont généralement les premiers théoriciens et praticiens de l'économie rustique :

« Il faut donc, lorsqu'on porte ses vûes sur les progrès de l'agriculture, voir beaucoup en détail & d'une manière suivie, la pratique des *fermiers* ; il faut souvent leur demander, plus souvent deviner les raisons qui les font agir. Quand on aura mis à cette étude le tems & l'attention nécessaires, on verra peut-être que la science de l'économie rustique est portée très-loin par les bons fermiers ; qu'elle n'en existe pas moins, parce qu'il y a beaucoup d'ignorans, mais qu'en général le courage & l'argent manquent plus que les lumières »<sup>45</sup>.

- 29 Associer l'économie rustique avec le fermage comme le fait Leroy dans FERMIER constitue un rejet explicite des pratiques économes et peu rentables des petits exploitants. La connaissance de la nature implique, chez Leroy et Quesnay – comme chez Linné<sup>46</sup> –, sa maîtrise. Dans l'article qui lui est consacré à la suite de celui de Rousseau, l'économie rustique est d'ailleurs définie comme

« l'art de connoître tous les objets utiles & lucratifs de la campagne, de se les procurer, de les conserver, & d'en tirer le plus grand avantage possible. Cette manière de s'enrichir est d'une étendue prodigieuse : c'est un tribut imposé sur tous les êtres de la nature ; les élémens même n'en sont pas exceptés. Ce seroit un ouvrage considérable que l'exposition seule des choses qui sont comprises dans l'économie rustique. Voici les principales. Celui qui vivra à la campagne, & qui voudra mettre son séjour à profit, connoitra l'agriculture & le jardinage dans tous les détails ; il n'ignorera rien de ce qui concerne les bâtimens nécessaires pour lui, pour sa famille, pour ses domestiques, pour ses animaux, & pour ses différentes récoltes ; la chasse, la pêche, la fauconnerie, les haras, les eaux, les forêts, les différens travaux rustiques ; plusieurs manufactures, telles celles de la fayence, de la poterie, de la chaux, de la brique, du fer, &c. Quelle que soit l'opinion vulgaire sur la vie d'un homme qui se livre tout entier à ces objets, je n'en connois aucune, sans exception, qui soit plus conforme à la nature, à la santé, à l'étendue des connoissances utiles, à l'élévation de l'esprit, à la simplicité des mœurs, au goût des bonnes choses, à la vertu, au bien public, à l'honnêteté & au bon sens »<sup>47</sup>.

- 30 L'article FROMENT (*économie rustique*) insiste également sur le fait que le froment « n'est point une production purement naturelle ; que l'existence de ce grain précieux n'est due qu'à la culture & à une longue suite de soins »<sup>48</sup>. Les planches d'économie rustique illustrent bien cet aspect de la perfection que l'homme donne à la nature à travers les représentations des procédés de la conservation du grain<sup>49</sup>, de 'l'art de faire éclore les poulets'<sup>50</sup>, ou du travail du coton, du tabac ou du sucre<sup>51</sup>.

- 31 L'économie rustique se présente donc dans l'*Encyclopédie* comme un mode de production intensive destinée à s'enrichir, à « tirer [de la nature] le plus grand avantage possible ». Par extension, l'économie de Quesnay peut être qualifiée de *politique* car elle applique les règles de la production agricole capitaliste à la société toute entière. L'une et l'autre se démarquent ouvertement de l'économie domestique incarnée par la petite exploitation familiale à vocation autarcique et peu exigeante en capitaux. Ceci constitue l'un des principaux points de rupture avec l'économie politique rousseausiste.

## b. La référence ambiguë de l'économie publique à l'économie domestique : la *prudens administratio* de Rousseau

- 32 Le rapport du gouvernement de la famille au gouvernement de l'État est au centre de la problématique politique aux XVII<sup>e</sup> et XVIII<sup>e</sup> siècles<sup>52</sup>. L'économique, ou gouvernement de la famille, représente un modèle de gouvernement politique. La scolastique semble notamment avoir largement contribué à établir une continuité théorique entre l'éthique, l'économique et la politique. Le père représente, en effet, l'autorité naturelle et religieuse par excellence<sup>53</sup>. Il est légitime dans son activité productive car celle-ci est associée à sa mission de conservation et de perpétuation de l'espèce.
- 33 La littérature politique, dite des *Miroirs*, va – par extension – associer le statut du roi à celui du père, d'abord pour des questions religieuses, puis pour des questions économiques<sup>54</sup>. Pour Montchrétien, le modèle paternel s'impose alors assez logiquement comme idéal politique<sup>55</sup>. L'*Economique du Prince* de François La Mothe Le Vayer, publié en 1653, s'inscrit dans la même démarche, et établit une continuité entre ce que l'on appellera plus tard l'économie domestique et l'économie politique. Le rapport entre ces deux types de société, que sont la famille et la Cité, est donc fondamental dans l'appréhension de l'économie politique au dix-huitième siècle que ce soit chez Voltaire<sup>56</sup>, Morellet<sup>57</sup>, ou chez Rousseau. La position de ce dernier est cependant originale, et a parfois été un peu rapidement réduite à une rupture de la tradition paternaliste.
- 34 Dans son article, Rousseau ne se propose pas d'interroger le principe, ni les différentes formes de la société politique – comme c'est le cas dans l'article de Boulanger –, mais son administration. Comment une société formée peut-elle remplir de manière pérenne son objet ? L'économie, c'est-à-dire selon la définition de Trévoux « l'ordre par lequel un corps politique subsiste principalement », permet de répondre à la question. Elle se compose, chez Aristote comme chez Rousseau, de deux rapports, le rapport aux personnes et le rapport aux biens. En ce sens, l'économie est synonyme de gouvernement. L'économie est, écrit Rousseau, « le sage et légitime gouvernement de la maison »<sup>58</sup> s'inspirant probablement lui aussi de Trévoux (le « sage et prudent gouvernement de la famille », *prudens administratio*).
- 35 Le terme d'économie, poursuit Rousseau, « a été dans la suite étendu au gouvernement de la grande famille, qui est l'état. Pour distinguer ces deux acceptions, on l'appelle dans ce dernier cas, *économie générale*, ou *politique* ; et dans l'autre cas, *économie domestique*, ou *particulière* »<sup>59</sup>. La substitution de la prudence par la légitimité du gouvernement est un point notable de la pensée de Rousseau. La dimension normative de l'économie se trouve toute entière concentrée dans la question politique de sa justice et de sa légitimité. C'est à ce niveau, nous semble-t-il, que l'on doit interroger le rapport de l'économie domestique à l'économie politique, autrement dit le statut du modèle de famille dans l'administration publique.
- 36 L'économie d'une famille s'assure spontanément par l'autorité naturelle exercée par le chef de famille, et l'intérêt commun qui unit les membres de la famille. En revanche, une société politique se caractérise par la convention et la divergence d'intérêt entre gouvernants et gouvernés. D'un côté la subsistance du corps constitue l'objectif commun des membres, de l'autre il se fait à leur détriment : « Le principal objet des travaux de toute la maison, est de conserver et d'accroître le patrimoine du pere, afin qu'il puisse un jour le partager entre ses enfans sans les appauvrir ; au lieu que la richesse du fisc n'est



qu'un moyen, souvent fort mal entendu, pour maintenir les particuliers dans la paix et dans l'abondance »<sup>60</sup>. Enfin, la pérennité se pose à des niveaux très différents pour l'une ou l'autre, « la petite famille est destinée à s'éteindre, et à se résoudre un jour en plusieurs autres familles semblables ; mais la grande [est] faite pour durer toujours dans le même état »<sup>61</sup>. Le gouvernement de la famille est naturel, il s'appuie sur l'inégalité physique et morale du chef. Il est par nature légitime.

- 37 L'objet de la société politique doit donc d'être de retrouver, par l'artifice, l'autorité et la légitimité naturelles du père de famille. Mais, pour gouverner l'État comme une famille, il faut que la nation devienne elle-même une 'grande famille', qu'elle transcende l'artificialité du lien social par la vertu civile. Rousseau distingue une économie publique légitime, pensée à l'image de l'économie domestique, d'une économie publique trompeusement inspirée du modèle domestique, celle du paternalisme politique<sup>62</sup>. L'idéal de l'économie politique entretient donc un rapport ambivalent avec l'économie domestique avec laquelle elle avait d'abord été opposée.
- 38 La légitimité de l'économie publique ou politique est fondée sur la vertu, soit la conformité des volontés particulières à la volonté générale. Elle donne à l'autorité politique une force comparable à l'autorité naturelle du père qui jouit d'une obéissance totale au sein de la famille sans avoir recours à la contrainte ou à la menace. La vertu d'un peuple, autrement dit sa citoyenneté, peut être développée par deux moyens : l'éducation publique et le patriotisme. La *patrie* constitue le cadre de référence dans lequel une communauté d'individus, tous mus par leur seul intérêt particulier, peut atteindre l'idéal politique que représente l'intérêt général. En cela, elle se *substitue* à la famille.
- 39 Toute la question de la légitimité et de la pérennité de la société se réduit donc à celle-ci : comment former des citoyens ? L'éducation des enfants est abandonnée « aux lumières et aux préjugés des pères » alors qu'elle « importe à l'état encore plus qu'aux pères »<sup>63</sup>. La mort du père interrompt de plus cette formation à un moment crucial de l'éducation civique, « l'état demeure, et la famille se dissout »<sup>64</sup>. Rousseau propose donc de transférer cette fonction paternelle à l'État et de subsister l'éducation publique à l'éducation domestique. Ce faisant, il transfère l'autorité paternelle à l'autorité politique. Désormais, le père se fera obéir de ses enfants par l'intermédiaire de la loi :
- « Que si l'autorité publique en prenant la place des pères, et se chargeant de cette importante fonction, acquiert leurs droits en remplissant leurs devoirs, ils ont d'autant moins sujet de s'en plaindre, qu'à cet égard ils ne font proprement que changer de nom, et qu'ils auront en commun, sous le nom de citoyens, la même autorité sur leurs enfants qu'ils exerçoient séparément sous le nom de *pères*, et n'en seront pas moins obéis en parlant au nom de la loi, qu'ils l'étoient en parlant au nom de la nature »<sup>65</sup>.
- 40 L'administration publique, la troisième règle de l'*économie publique*, peut alors légitimement s'inspirer de l'*économie domestique* : « Ainsi après avoir parlé de l'*économie générale* par rapport au gouvernement des personnes, il nous reste à la considérer par rapport à l'administration des biens »<sup>66</sup>. L'économie publique a pour objet essentiel l'impôt et son mode de perception<sup>67</sup>. Celle-ci est au cœur de la question économique, c'est-à-dire de la gestion prudente des revenus et des dépenses, mais également de la nature du contrat social. Le dilemme qui se pose à Rousseau est donc le suivant : comment un État peut-il subvenir à ses besoins sans atteindre à la propriété privée de ses membres ? Comment assurer le bon fonctionnement de l'État en taxant le moins possible et le plus justement possible ? Par « une juste et sage *économie* »<sup>68</sup>. La nation, devenue

'grande famille' doit être gérée comme telle, c'est-à-dire avec économie, en bon père de famille :

« De cette règle découle la plus importante maxime de l'administration des finances, qui est de travailler avec beaucoup plus de soin à prévenir les besoins, qu'à augmenter les revenus [...]. Je crois que de cette grande maxime bien établie, découloient les prodiges des gouvernements anciens, qui faisoient plus avec leur parcimonie, que les nôtres avec tous leurs trésors ; et c'est peut-être de-là qu'est dérivée l'acception vulgaire du mot *économie*, qui s'entend plutôt du sage ménagement de ce qu'on a, que des moyens d'acquérir ce qu'on a pas »<sup>69</sup>.

- 41 Rousseau fait de l'*économie domestique*, au sens restreint du terme, i.e. la gestion du patrimoine familial, le modèle de l'administration politique. La nation peut ainsi, comme la famille, se réunir autour d'un patrimoine commun. Idéalement, l'État devrait tirer ses revenus d'un domaine foncier public dont la gestion constitue aux yeux de Rousseau « le vrai système économique »<sup>70</sup>. L'économie, ici entendue dans son acception moderne, est donc l'un des vecteurs du gouvernement légitime. En faisant de l'administration domestique un principe politique, l'économie, entendue comme *prudens administratio*, contribue à la citoyenneté de ses membres et à l'autorité de la loi. Le système commercial, fondé sur une division poussée du travail, une circulation accrue des marchandises et de la monnaie, est en grande partie responsable, selon Rousseau, de la perversion des mœurs (vanité, amour-propre) et de la décadence des civilisations occidentales. La société politique se réunit autour de la loi, alors qu'elle se déchire autour des richesses<sup>71</sup>.
- 42 Pourquoi aller au-delà de sa discipline pour mieux en faire l'histoire ? Nous avons débuté cet article par une étude lexicographique, nous l'achevons par une comparaison entre Quesnay et Rousseau. En quoi ce détour a-t-il été nécessaire ? Il l'a été, croyons nous, au moins à deux niveaux.
- 43 Tout d'abord, les contributions de Quesnay et Rousseau étaient, par nature, lexicographiques. L'une et l'autre ont participé, à des niveaux différents, à la diffusion du syntagme *économie politique* dans le langage usuel. En apprécier la valeur imposait une comparaison avec les dictionnaires utilisés à l'époque. L'une et l'autre appartiennent, de plus, à une structure encyclopédique élaborée composée d'articles, de sous-entrées, de désignants et de renvois. Le niveau de l'une et de l'autre, au sein du système encyclopédique, doit alors être clairement identifié. Paradoxalement, il apparaît que ni Quesnay ni Rousseau n'ont explicitement choisi *économie politique* pour désigner leur écrit. Elle leur a été fortuitement accolée, dans un cas par confusion avec la réédition de l'article sous forme de brochure, dans l'autre au moyen du désignant.
- 44 Il était ensuite nécessaire de rétablir l'articulation entre les différentes branches de l'*économie*, notamment l'*économie rustique* et l'*économie domestique* toutes deux reconnues dans le système encyclopédique, et la nouvelle *économie politique*. La généalogie de l'économie politique ou des économies politiques, entendues comme champs du savoir, ne peut, en effet, surgir du néant. Il n'y a, au milieu du dix-huitième siècle, aucune identification commune d'une branche du savoir dénommée *économie politique*. Rousseau et Quesnay ont contribué à donner un contenu à cette désignation, qui avait jusqu'alors autant de significations différentes que l'*économie* en avait elle-même. Mais leurs perspectives sont très différentes. Alors que Rousseau propose une réflexion sur le modèle *domestique* du gouvernement, i.e. de l'*économie*, politique, Quesnay envisage les effets d'une *économie rustique* à l'échelle nationale. L'un et l'autre se positionnent à des niveaux sémantiques totalement différents. Seule cette expression, qu'ils ne revendiquent pas, mais que l'histoire leur a attribuée, semble les réunir ici. Cependant, une position



théorique les réunit effectivement, ou plus exactement les rapproche en les opposant indirectement : l'articulation interne entre l'*économie domestique*, au sens rousseauiste du terme, et l'*économie rustique*. Le modèle offert par l'*économie domestique* aristotélicienne, caractérisée par une production d'autosuffisance, un faible niveau d'échange et une utilisation raisonnée de la monnaie, s'oppose, en effet, presque frontalement à celui de l'*économie rustique* sur lequel s'appuie Quesnay et qu'on appellera par la suite, et de manière définitive, l'*économie politique*.

---

## NOTES

1. M.F. Piguet, « Œconomie/Économie (politique) dans le texte informatisé de l'Encyclopédie », *Recherches sur Diderot et sur l'Encyclopédie*, p. 31-32, Avril 2002, pp. 123-137.
2. *Dictionnaire universel contenant tous les mots françois tant vieux que modernes, & les termes de toutes les sciences et des arts* [par A. Furetière], [1690], tome I, [non paginé].
3. « ŒCONOMIQUE. Adj. Qui appartient à l'oeconomie. Aristote a écrit deux livres de l'*Oeconomique*. Les maximes *oeconomiques* changent selon les occasions », *Dictionnaire universel*, *op. cit.*, [1690, np].
4. M. Leca-Tsiomis, *Ecrire l'Encyclopédie. Diderot : de l'usage des dictionnaires à la grammaire philosophique*, Oxford, Voltaire Foundation, 1999.
5. *Dictionnaire universel françois et latin* [dit dictionnaire de Trévoux], 5<sup>e</sup> éd., 1752, III, 494. Cette définition sera légèrement modifiée dans la sixième édition en « sage et prudent gouvernement d'une maison ; [la] règle qu'on apporte dans la conduite d'un ménage, dans la dépense d'une maison », *Trévoux*, 6<sup>e</sup> éd., 1771, p. 555.
6. « Dans cette acception, le mot d'*Economie* est relatif à l'usage qu'on fait de son bien : c'est une juste dispensation du bien que l'on a ; un emploi convenable de ses fonds, un moyen industrieux de les perpétuer, pour être toujours à portée de ne pas diminuer sa dépense, & même l'augmenter, en multipliant sans interruption le produit des sommes qu'on fait circuler avec honneur. Son grand art est de tirer parti de tout ce qui est entre ses mains, & de ne rien dissiper. Un prudent père de famille accroît ses biens par une prévoyante *économie*. », *Trévoux*, 6<sup>e</sup> éd., 1771, p. 555.
7. *Trévoux*, 5<sup>e</sup> éd., 1752, III, 494.
8. « *Legalis*. C'est la manière dont Dieu jugea à propos de conduire son peuple, par le ministère de Moïse. Elle comprenoit non-seulement les loix politiques & cérémoniales, mais aussi la loi morale, entant qu'elle prononçoit malédiction contre tous ceux qui ne l'accompliroient pas parfaitement », *Trévoux*, 5<sup>e</sup> éd., 1752, III, 495.
9. « *Evangelica*, se dit par opposition à l'*économie* légale, & renferme tout ce qui appartient à l'alliance de grace que Dieu a traitées avec les hommes par JÉSUS-CHRIST », *Trévoux*, 5<sup>e</sup> éd., 1752, III, 495.
10. « les Théologiens distinguent deux *oeconomies*, l'ancienne & la nouvelle, ou, pour m'exprimer en d'autres termes, l'*oeconomie* légale & l'*oeconomie* évangélique ; l'*oeconomie* légale est celle du ministère de Moïse, qui comprend les lois politiques & cérémonielles du peuple juif ; l'*oeconomie* évangélique, c'est le ministère de Jesus Christ, sa vie & ses préceptes. », *ŒCONOMIE, Enc.*, XI, 360a.
11. *Trévoux*, 6<sup>e</sup> éd., 1771, p. 556.

12. *Dictionnaire de l'Académie française*, Paris, 1718, T. II., p. 156.
13. *Trévoux*, 5<sup>e</sup> éd., 1752, III, 495.
14. « C'est l'art de tirer le plus grand avantage possible des biens de la campagne », *Trévoux*, 6<sup>e</sup> éd., 1771, p. 555. L'entrée ÉCONOMIE RUSTIQUE de l'*Encyclopédie* débute ainsi : « l'art de connoître tous les objets utiles & lucratifs de la campagne, de se les procurer, de les conserver, & d'en tirer le plus grand avantage possible », ÉCONOMIE RUSTIQUE, *Enc.*, V, 349a.
15. « Termes impropres, dont on se sert quelquefois pour désigner l'animal même. De-là ces façons de parler abusives, mouvemens, fonctions de l'économie animale. À parler exactement, cette dénomination ne regarde que le mécanisme, l'ordre, l'ensemble des fonctions & des mouvemens qui entretiennent la vie des animaux, dont l'exercice parfait constitue l'état de santé, dont le moindre dérangement est par lui-même maladie, & dont la cessation est la mort. C'est dans ce sens qu'on dit, mouvemens, lois de l'économie animale », *Trévoux*, 6<sup>e</sup> éd., 1771, p. 556. On trouve au début de l'article de Ménuret de Chambaud dans l'*Encyclopédie* le passage suivant : « quelques auteurs ont employé improprement le nom d'*oeconomie animale*, pour désigner l'animal lui même ; c'est de cette idée que sont venues ces façons de parler abusives, *mouvemens, fonctions de l'oeconomie animale* ; mais cette dénomination prise dans le sens le plus exact & le plus usité ne regarde que l'ordre, le mécanisme, l'ensemble des fonctions & des mouvemens qui entretiennent la vie des animaux, dont l'exercice parfait, universel, fait avec constance, alacrité & facilité, constitue l'état le plus florissant de santé, dont le moindre dérangement est par lui-même *maladie*, & dont l'entière cessation est l'extrême diamétralement opposé à la vie, c'est-à-dire la mort. L'usage, maître souverain de la diction, ayant consacré cette signification, a par-là même autorisé ces expressions usitées, *lois de l'oeconomie animale, phénomènes de l'oeconomie animale*, qui sans cela & suivant l'étymologie présenteroient un sens absurde, & seroient un pléonasme ridicule. », ÉCONOMIE ANIMALE, *Enc.*, XI, 360a. La thèse défendue par Ménuret de Chambaud, ainsi que par Bordeu dans l'article CRISE, est celle du vitalisme auquel Diderot souscrivit en partie. Voir D. Boury, « Théophile de Bordeu : source et personnage du *Rêve de D'Alembert* », *RDE*, 34, 2003, pp. 11-24 ; C. Duflo, « Diderot et Ménuret de Chambaud », *RDE*, 34, 2003, pp. 25-44 et R. Rey, *Naissance et développement du vitalisme en France*, Oxford : Voltaire Foundation, 2000.
16. L'édition de 1740 du *Dictionnaire de l'Académie* ne mentionne cependant ni d'économie rustique ni d'économie animale.
17. Voir à ce propos M. Leca-Tsiomis, *Ecrire l'Encyclopédie*, ouvr. cité.
18. OECONOMIE, *Enc.*, XI, 359b.
19. De nombreux commentateurs de Rousseau ont, par le passé, opéré un raccourci malencontreux entre l'article *Économie* et le *Discours sur l'économie politique*. Anne Eyssidieux-Vaissermann, qui consacre pourtant son étude à l'article de Rousseau, le désigne systématiquement par ÉCONOMIE POLITIQUE : « Il y a en réalité dans l'*Encyclopédie* deux articles portant un titre presque identique : l'article de Rousseau : ÉCONOMIE POLITIQUE et l'article de Boulanger : OECONOMIE POLITIQUE, publié dans le t. IX [sic] en 1765 », A. Eyssidieux-Vaissermann, « Rousseau et la science de l'économie politique dans l'*Encyclopédie* », *Kairos*, n° 18, *Ordre et production des savoirs dans l'Encyclopédie de Diderot et de d'Alembert*, 2001, p. 50.
20. « L'auteur de l'article ÉCONOMIE POLITIQUE de ce Dictionnaire est de même sentiment quant à la nature de l'impôt ; mais il ne veut pas qu'il soit payé par le marchand, & prétend qu'il doit l'être par l'acheteur. J'avoue que je ne vois dans cette différence que des chaînes ajoutées à la liberté des citoyens, & une contradiction de plus dans celui qui s'en dit le plus grand défenseur. Néron ne fit qu'ordonner l'inverse de ce que propose M. Rousseau, & parut, dit Tacite, avoir supprimé l'impôt. C'étoit celui de quatre pour cent, qu'on le voit sur le prix de la vente des esclaves. Tant il est vrai que la forme y fait quelque chose, & que celle du citoyen de Genève n'est pas la meilleure. », VINGTIÈME, *Enc.*, XVII, 868a-b. L'entrée OECONOMIE signale également l'existence d'une 'oeconomie politique' dont il est traité dans l'article de Rousseau.

21. ÉCONOMIE, *Enc.*, V, 337a.
22. On doit à l'éditeur Jacob Vernes l'initiative de rééditer cet article sous le titre de *Discours sur l'économie politique* « avec l'accord de Rousseau, bien que cette édition ait été réalisée hors de son contrôle », B. Bernardi, Introduction, in J.-J. Rousseau, *Discours sur l'économie politique*, Paris : Vrin, Collection Textes & Commentaires, 2002, p. 8.
23. Il a été souvent dit que cette publication constituait une édition largement remaniée par le baron d'Holbach. Paul Sadrin a montré, en comparant les versions manuscrites et l'édition du baron d'Holbach, que ce n'était pas le cas. P. Sadrin, *Nicolas-Antoine Boulanger (1722-1759) ou avant nous le déluge*, Oxford : The Voltaire Foundation, 1986, p. 45.
24. « Toute l'économie de la société humaine est appuyée sur ce principe général et simple : je veux être heureux ; mais je vis avec des hommes qui, comme moi, veulent être heureux également chacun de leur côté : cherchons le moyen de procurer notre bonheur, en procurant le leur, ou du moins sans jamais y nuire », SOCIÉTÉ, *Enc.*, XV, 252b.
25. « C'est sans doute à la suite de ces tems déplorables qui avoient réduit l'espèce humaine, renversé son séjour, & détruit sa subsistance, qu'ont dû être faits ces réglemens admirables, que nous trouvons chez les anciens peuples, sur l'agriculture, sur le travail & l'industrie, sur la population, sur l'éducation, & sur tout ce qui concerne l'oeconomie civile et domestique » (*Despotisme oriental*, 1761, p. 48) devient « C'est sans doute à la suite de tous les événemens malheureux qui ont autrefois ruiné l'espece humaine, son séjour & sa subsistance, qu'ont dû être faits tous ces réglemens admirables, que nous retrouvons chez les peuples les plus anciens, sur l'agriculture, sur le travail, sur l'industrie, sur la population, sur l'éducation, & sur tout ce qui concerne l'œconomie publique et domestique. », ŒCONOMIE POLITIQUE, *Enc.* XI, 367b-368a.
26. Dans une lettre du 20 novembre 1755 Boulanger écrit à propos du *Despotisme oriental* : « les révolutions de la nature faisant la baze de mes principes, c'est la que j'ay toujours été chercher l'homme ; et vous verrés icy que pour raisonner sur les premieres societés je n'ay point été consulter l'homme sauvage, ni l'homme dans l'État de nature, animal ou corrompu, ni enfin l'homme métaphisique, mais l'homme échapé aux malheurs du monde ; j'ai été chercher un Etre reel, que j'ai étudié dans un État reel », P. Sadrin, *Nicolas-Antoine Boulanger, op. cit.*, p. 238. On trouve un passage similaire dans ŒCONOMIE POLITIQUE, *Enc.*, XI, 367a.
27. P. Sadrin, *Nicolas-Antoine Boulanger, op. cit.*, pp. 51-52.
28. Voir à ce sujet, M. Minerbi, « L'économie politique des Anciens et celle des Modernes dans l'*Encyclopédie* », *Recherches sur Diderot et sur l'Encyclopédie*, p. 12, Avril 1992, pp. 34-39.
29. M.F. Piguet, « Œconomie/Économie (politique) dans le texte informatisé de l'*Encyclopédie* », *Recherches sur Diderot et sur l'Encyclopédie*, p. 31-32, Avril 2002, pp. 123-137.
30. Je remercie Marie-France Piguet pour cette information.
31. J. Proust, *Diderot et l'Encyclopédie*, Paris : Albin Michel, 1995, pp. 487-491.
32. « RASP-HUIS, (*Hist. mod. Économie politiq.*) c'est ainsi que l'on nomme à Amsterdam, & dans d'autres villes de la province de Hollande, des maisons de correction, dans lesquelles on enferme les mauvais sujets, les vagabonds & gens sans aveu, qui ont commis des crimes pour lesquels les lois n'ont point décerné la peine de mort. On occupe les prisonniers à des travaux pénibles, au profit du gouvernement. À Amsterdam le principal de ces travaux consiste à raper des bois des Indes fort durs, pour servir dans les teintures ; c'est-là ce qui fait appeler ces sortes de maisons de force rasp-huis, ce qui signifie maison où l'on rape », RASP-HUIS, *Enc.*, XIII, 816a.
33. « SPINHUYS, s.m. (*Hist. mod. Econom. politique*) ce mot est hollandois, & signifie maison où l'on file ; on donne ce nom en Hollande à des maisons de force établies dans presque toutes les villes, dans lesquelles on renferme les femmes de mauvaise vie, qui ont attiré l'attention de la police ; on les y occupe à filer & à différens autres travaux convenables à leur sexe ; on leur épargne point les corrections, lorsqu'elles manquent à remplir la tâche qui leur est imposée. Ces sortes de maisons sont ordinairement sous la direction de deux échevins, qui nomment un inspecteur & une inspectrice, qui leur rendent compte », SPINHUYS, *Enc.*, XV, 463a.

34. Le désignant d'*économie rustique* n'a pas été retenu probablement en raison des considérations d'économie publique développées dans l'article. Notons également dans celui-ci l'emploi de l'expression *économie forestière* comme subdivision de l'économie rustique.

35. Voir, par exemple, P. Steiner, *La « science nouvelle » de l'économie politique*, Paris, P.U.F., 1998, ainsi que les actes du colloque *Commerce, population et société autour de Vincent de Gournay (1748-1758)*, Paris, 19-21 février 2004, à paraître prochainement aux éditions de l'INED.

36. Tous les esprits se sont focalisés sur le *Traité d'économie politique* d'Antoine de Montchrétien publié en 1615. On sait aujourd'hui que les expressions d'économie publique ou politique lui sont bien antérieures. Jacques Lefèvre d'Étaples dans l'édition latine qu'il donne en 1506 des *Economiques*, alors attribuées à Aristote, intitule le livre II 'Économie publique' (*Oeconomiarum publicarum Aristotelis liber unus*). Louis Turquet de Mayerne est, dans l'état actuel des connaissances, le premier à utiliser 'économie politique' dans un écrit imprimé français, *La Monarchie aristodémocratique, ou le gouvernement composé et meslé des trois formes de légitimes républiques*, rédigé vers 1590 mais publié seulement en 1611. Voir, A. Béraud et G. Faccarello (eds), *Nouvelle histoire de la pensée économique*, Paris : Éditions La Découverte, tome 1, 1992, p. 21.

37. « Il y a dans l'oeconomie animale, comme dans l'oeconomie politique, un enchaînement de maux qui naissent d'un premier vice dans le principe, dont la force entraîne tout », *DYSPEPSIE, Enc.*, V, 177a.

38. Cette hypothèse aurait pu être défendue si *FERMIERS* avait été publié après le *Discours sur l'économie politique* de Rousseau, mais le volume VI de l'*Encyclopédie* parut en 1756. La confusion dont nous avons parlé précédemment n'avait donc pas lieu d'être.

39. L'article *FERME* est également cité par Quesnay dans *HOMMES*, article non publié dans l'*Encyclopédie*, et sans désignant.

40. « on imagine que les travaux du pauvre cultivateur sont aussi avantageux que ceux du riche fermier », « Il s'en faut de peu qu'on ne croie que l'usage des chevaux et l'usage des bœufs ne soient également avantageux », « On croit vulgairement que les bœufs ont plus de force que les chevaux », « On croit vulgairement qu'il y a beaucoup plus de profit, par rapport à la dépense, à labourer avec des bœufs, qu'avec des chevaux », « On dit que les chevaux sont plus sujets aux accidents et aux maladies que les bœufs », « On a cru que la politique regardait l'indigence des habitants de la campagne, comme un aiguillon nécessaire pour les exciter au travail », « On a prétendu qu'il fallait restreindre la culture du blé », « Les habitants de la ville croient ingénument que ce sont les bras des paysans qui cultivent la terre »....

41. Il appuie son raisonnement sur des données quantitatives précises, le nombre de chevaux et de bœufs nécessaires par arpent cultivé, le prix des chevaux et des bœufs, le coût d'opportunité de leur achat, l'assolement pratiqué dans un cas comme dans l'autre, une analyse comparative de leurs besoins, de leur production...

42. « Épargne, dans le sens le plus vulgaire, est une dépendance de l'économie ; c'est proprement le soin & l'habileté nécessaires pour éviter les dépenses superflues, & pour faire à peu de frais celles qui sont indispensables. Les réflexions que l'on va lire ici, auroient pû entrer au mot ÉCONOMIE, qui a un sens plus étendu, & qui embrasse tous les moyens légitimes, tous les soins nécessaires pour conserver et accroître un bien quelconque, & sur-tout pour le dispenser à propos. C'est en ce sens que l'on dit économie d'une famille, économie des abeilles, économie nationale » (*ÉPARGNE, Enc.*, V, 745a).

43. *FERMIER, Enc.*, VI, 528a.

44. *FERME, Enc.*, VI, 511a.

45. *FERMIER, Enc.*, VI, 528b.

46. *Principes de l'Economie, fondés sur la Science naturelle & sur la Physique*. Par M. Linnæus, Docteur en Médecine, de l'Académie des Sciences de Stokolm.

47. *ÉCONOMIE RUSTIQUE, Enc.*, V, 349a

48. *FROMENT, Enc.*, VII, 334b.

49. Toutes les planches de l'*Encyclopédie Diderot et D'Alembert*, Paris : EDDL, 2001, p. 27.
50. *Idem*, pp. 32-33.
51. *Idem*, pp. 41-45.
52. M. Foucault, *Sécurité, territoire, population. Cours au Collège de France. 1977-1978*, Paris, Seuil-Gallimard, pp. 91-118.
53. Sur le statut politique et religieux de la parenté au Moyen-âge, voir notamment J. Baschet, *La civilisation féodale. De l'an mil à la colonisation de l'Amérique*, Paris, Aubier, 2004, pp. 426-459.
54. M. Senellart, *Les arts de gouverner. Du régime médiéval au concept de gouvernement*, Paris, Seuil, 1995. Voir aussi G. Duby, *Les trois ordres ou l'imaginaire du féodalisme*, réédité in *Féodalité*, Paris : Quarto Gallimard, 1996, pp. 472-477. J. Le Goff, *Saint Louis*, réédité in *Héros du Moyen Age, le Saint et le Roi*, Paris, Quarto Gallimard, 2004, pp. 515-539.
55. « Les vacations [professions] privées font la publique. La maison est premier que la Cité. La ville que la Province. La Province que le Royaume. Ainsi l'art politic dépend médiatement de l'oeconomie. Et comme il en tient beaucoup de conformité, il doit pareillement emprunter son exemple. Car le bon gouvernement domestic, à le bien prendre, est un patron et modèle du public ; soit que l'on regarde le droit commandement, soit la fidèle obéissance, liaison principale de l'un et de l'autre. » (A. de Montchrétien, *Traicté de l'Economie politique*, Genève, Librairie Droz, 1999, p. 52). Funck-Brentano republia en 1889 le *Traité* de Montchrétien avec une longue introduction intitulée « L'économie politique patronale ».
56. Dans le *Dictionnaire philosophique*, Voltaire définit l'économie comme l'administration d'un bien. Celle-ci est commune au père de famille et au surintendant des finances. Il ne distingue donc pas les principes de l'économie domestique, qu'il associe également à l'économie rustique, et de l'économie publique.
57. C. Salvat, « 'De la science de l'économie publique' de Morellet », in *Studies on Voltaire and the Eighteenth Century*, 2003, 10 pp. 65-95.
58. ÉCONOMIE, *Enc.*, V, 337a.
59. ÉCONOMIE, *Enc.*, V, 337a.
60. ÉCONOMIE, *Enc.*, V, 337b.
61. ÉCONOMIE, *Enc.*, V, 337b.
62. De manière analogue, pour Boulanger, les lois civiles établies à l'image des lois domestiques ne permettent pas de répondre aux spécificités des sociétés politiques. Parce que les hommes ont fait de la coutume et de l'histoire une légitimité en soi, au détriment de la raison même, l'âge d'or s'est progressivement et insensiblement transformé en âge de plomb : « À mesure que le séjour de l'homme s'est embelli, à mesure que les sociétés se sont multipliées, & qu'elles ont formé des villes & des états, le regne moral a dû nécessairement faire place au regne politique, & le tien & le mien ont dû paroître dans le monde, non d'abord d'homme à homme, mais de famille à famille & de société à société, parce qu'ils y sont devenus indispensables, & qu'ils font partie de cette même harmonie qui a dû rentrer parmi les nations renouvelées, comme elle est insensiblement rentrée dans la nature après le dernier chaos. Cet âge d'or a donc été un état de sainteté, un état surnaturel digne de notre envie, & qui a justement mérité tous les regrets de l'antiquité : cependant lorsque les législations postérieures en ont voulu adopter les usages & les principes sans discernement, le bien s'est nécessairement changé en mal, & l'or en plomb », ÉCONOMIE POLITIQUE, *Enc.*, XI, 368a.
63. ÉCONOMIE, *Enc.*, V, 343b.
64. ÉCONOMIE, *Enc.*, V, 343b.
65. ÉCONOMIE, *Enc.*, V, 343b.
66. ÉCONOMIE, *Enc.*, V, 344a.
67. Il existe alors deux systèmes de perception alternatifs, la régie et la ferme défendues respectivement par Rousseau et Quesnay. Dans les articles 'Fermes du Roi' (*Finances*) et 'Financier' (*Politiq.*), Pesselier critique sévèrement Montesquieu pour ses attaques contre la ferme

et les financiers. Il reprend, dans ce premier article, un à un les huit principes du philosophe bordelais défendant la régie, dont le premier : « La *régie* est l'administration d'un bon père de famille, qui leve lui-même avec *économie & ordre* ses revenus » (FERMES, *Enc.*, VI, 513b). Un contributeur inconnu lui répondra dans l'article SUBSIDES.

68. « si les contributions sont volontaires, elles ne produisent rien ; si elles sont forcées, elles sont illégitimes ; et c'est dans cette cruelle alternative de laisser périr l'état ou d'attaquer le droit sacré de propriété, qui en est le soutien, que consiste la difficulté d'une juste et sage économie », ÉCONOMIE, *Enc.*, V, 344d.

69. ÉCONOMIE, *Enc.*, V, 345a.

70. ÉCONOMIE, *Enc.*, V, 346a.

71. « La puissance civile s'exerce de deux manières : l'une légitime par l'autorité, l'autre abusive par les richesses. Partout où les richesses dominant, la puissance et l'autorité sont ordinairement séparées, parce que les moyens d'acquérir la richesse et les moyens de parvenir à l'autorité n'étant pas les mêmes sont rarement employés par les mêmes gens. Alors que la puissance apparente est dans les mains des magistrats et la puissance réelle est dans celle des riches. Dans un tel gouvernement tout marche au gré des passions des hommes, rien ne tend au but de l'institution » J.-J. Rousseau, *Projet de Constitution pour la Corse, Œuvres Complètes*, Paris, Gallimard, 1964, III, 939.

## RÉSUMÉS

Le regard de l'historien de la pensée économique sur l'*Encyclopédie* est, le plus souvent, attiré par les deux contributions séminales, que sont les articles FERMIERS et GRAINS de Quesnay, considéré comme l'un des fondateurs de la discipline. La curiosité est également attisée par un événement linguistique majeur, l'usage jusque-là rarissime de la dénomination 'moderne' d'*économie politique*, comme désignant de ces deux articles. L'affaire semblait claire, l'économie politique, en tant que discipline nouvelle, serait née dans l'*Encyclopédie*. Mais les travaux récents de M.F. Piguet ont montré la complexité inhérente aux dénominations économiques dans l'*Encyclopédie* et les dangers d'une démarche réductrice. Il faut donc interroger préalablement œ/économie et ses composés : économie rustique, animale, domestique, politique. Nous inscrivant dans la continuité de ces travaux, nous tentons plus spécifiquement d'identifier deux sémantiques de l'économie politique, articulées chacune autour d'un syntagme lexicalement identifié, l'économie domestique et l'économie rustique.

### Articles 'Œ/Economie' and their subject indicators in the *Encyclopédie*

When historians of economic thought have looked at the *Encyclopédie*, they have been mainly interested in two seminal articles (FERMIERS and GRAINS) by Quesnay, considered as one of the founders of the discipline. Their curiosity is also aroused by the major linguistic event constituted by the previously extremely rare 'modern' denomination of 'économie politique', given as the subject indicator of these two articles. They concluded that the new discipline of economics was born in the *Encyclopédie*. However, recent work by M.F. Piguet has shown the complexity inherent in the work's economic denominations and the dangers of oversimplification. We first need to look at œ/économie and its combinations, namely rural, animal, domestic and political economy. This article, following on from this research, tries to

identify specifically two semantics of political economy, each relating to a lexically identified syntagma, namely domestic economy and rustic economy.

AUTEUR

**CHRISTOPHE SALVAT**

CNRS

