

Le mérite
Signification, possibilité et valeur

Aurélien Allard

Thèse de doctorat
Discipline : Philosophie

Directeur de thèse :
Bertrand Guillaume, Université Paris 8

Université Paris 8

Table des matières

Remerciements	3
Introduction	5
Première partie Fondations conceptuelles de l'idée de mérite	14
Chapitre 1 Définir le mérite.....	15
Chapitre 2 Les critères du mérite.....	43
Chapitre 3 La centralité du mérite dans les conceptions morales ordinaires.....	60
Deuxième partie De l'importance des données empiriques pour une philosophie normative. 87	
Chapitre 4 Pourquoi une philosophie expérimentale ?.....	88
Chapitre 5 La question de l'expertise philosophique	122
Troisième partie Justifier le mérite.....	144
Chapitre 6 Justifier le mérite : Trois premières approches.....	145
Chapitre 7 Justifier le mérite par sa valeur communicative : Introduction aux débats théoriques	174
Chapitre 8 La valeur communicative de la peine : données empiriques	186
Chapitre 9 Justifier les théories communicationnelles de la peine	218
Quatrième partie Mérite et responsabilité	230
Chapitre 10 Responsabilité et égalitarisme de la fortune	231
Chapitre 11 Responsabilité et mérite : études expérimentales	244
Chapitre 12 L'importance de ce dont nous ne sommes pas responsables	262
Conclusion Défense du mérite et théorie pluraliste de la justice.....	272
Appendice général Méthodes expérimentales et problèmes de généralisation	284
Bibliographie.....	297

Remerciements

J'ai bénéficié, tout au long de cette thèse, du soutien d'un très grand nombre d'amis, de proches, et de collègues.

Je tiens en tout premier lieu à remercier mon directeur de thèse, Bertrand Guillarme, pour son soutien tout au long de cette thèse, et pour ses remarques et commentaires extrêmement pertinents sur les premières versions de ce travail.

Même si leur influence sur cette thèse a été moins directe, je souhaiterais également remercier quatre philosophes qui ont eu une grande influence sur mon orientation philosophique. Tout d'abord, je souhaiterais remercier Chantal Hasnaoui, ma professeure de khâgne, qui m'a véritablement initié à la philosophie et changé mon regard sur le monde. Plus tard, j'ai également bénéficié à l'École normale supérieure d'un tutorat extrêmement bénéfique de la part de Mathias Girel, qui a su me guider pendant tout mon temps à l'école, et grâce à qui j'ai pu découvrir toutes les richesses de la philosophie américaine. Toujours à l'ENS, en troisième année de licence, j'ai eu la chance de faire un stage de philosophie expérimentale auprès de Florian Cova. Ça a été une expérience fantastique, qui a été ma première véritable introduction au monde de l'expérimentation. Enfin, je tiens à remercier Sophie Guérard de Latour, qui a dirigé mes travaux de mémoire de Master avec beaucoup d'attention. Le deuxième mémoire portait déjà sur le thème du mérite, et je lui suis encore reconnaissant pour toutes les remarques et critiques qu'elle y a porté.

Cette thèse a été élaborée sur deux continents, et dans trois pays. J'ai effectué deux séjours de recherche à Belfast, auprès de Paulo Sousa, à l'automne 2015 et 2016. La collaboration avec Paulo a été extrêmement agréable et utile. Je remercie également Samuel Arty Ward, Hugh Turpin, et Mathilde Hernu, pour les nombreuses discussions de psychologie morale que nous avons eues à l'époque.

Je suis également resté un an à Philadelphie, pour travailler avec Geoff Goodwin sur des expériences de psychologie morale. Là encore, l'expérience a été très riche, et ça a été un immense plaisir de travailler avec Geoff. Je tiens aussi à remercier Hanne M. Watkins et Ivan

Denker pour de multiples discussions sur la psychologie morale et sur l'amélioration des méthodes en psychologie. L'année à Philly aurait été beaucoup moins agréable sans eux.

Sur Paris, je voudrais remercier tout particulièrement les membres de *Sodalicum*, et notamment Antoine Marie, Sacha Yesilaltay, Joffrey Fuhrer, Charlotte Barot, Fabien Dézèque, Léo Fitouchi, Matthias Michel, Benoît de Courson, Loïa Lamarque, et Camille Williams pour toutes les discussions que nous avons eues au cours de la dernière année de thèse. *Sodalicum* a représenté un soutien moral énorme au cours de cette période.

Enfin, l'aide apportée par ma famille a été fantastique tout au long de cette thèse. Je tiens à remercier mes parents, ma sœur, mon frère, pour tout le temps passé ensemble pendant ces trois ans, et pour ces voyages irlandais ou américains que nous avons faits ensemble. Je tiens également tout particulièrement à remercier ma sœur, qui a relu un chapitre de la thèse, mon père, qui en a relu quelques-uns, et ma mère, qui a relu la quasi-intégralité du travail. L'ensemble du texte serait beaucoup moins lisible sans le temps qu'ils ont pris à me relire et à me corriger.

Introduction

Le mérite constitue une notion fondamentale dans la vie courante et dans le débat politique. On retrouve constamment des références au mérite dans les revendications de justice sociale. L'État devrait soutenir les étudiants méritants. Un bon employé devrait recevoir le salaire qu'il mérite. Les promotions devraient être reconnues selon les mérites des individus, et non selon le genre ou l'origine ethnique. Dans le domaine pénal, l'idée de mérite semble également fondamentale. Seul le coupable mérite d'être puni. Un criminel mérite de recevoir une punition proportionnée, ni trop peu ni trop sévère.

Il est difficile de s'imaginer à quoi ressemblerait notre vie morale si elle était expurgée de l'idée de mérite. De fait, de multiples enquêtes internationales ont constaté sa popularité dans un très grand nombre de pays. Dire que quelqu'un mérite de recevoir quelque chose semble une raison très forte de la lui donner. On pourrait même dire que le mérite semble l'idée la plus fondamentale pour constituer une théorie de la justice, voire qu'une théorie de la justice est avant tout une théorie du mérite.

Les remises en cause de l'idée de mérite

Pourtant, malgré leur caractère extrêmement courant et banal, les jugements de mérite cachent de fortes difficultés. D'un point de vue purement philosophique, il est possible de remettre en cause la validité du concept de mérite, de douter que l'idée courante de mérite possède un véritable sens. D'un point de vue proprement politique, il est également possible de remettre en cause la valeur du mérite, en ce que la notion semble remettre en cause l'idée d'égalité. A tout le moins, les deux notions semblent entrer dans de possibles conflits.

Un premier reproche possible à la notion de mérite est qu'elle semble être terriblement vague. On pourrait en effet soutenir que le fait qu'un individu mérite d'obtenir telle ou telle

récompense est plus ou moins synonyme de dire qu'il est bon ou approprié de lui fournir cette récompense. Plus généralement, on peut dire par exemple que tout être humain mérite d'être heureux, que tout le monde mérite de voir ses besoins satisfaits, etc. Il devient alors difficile d'envisager que l'idée de mérite ait un contenu spécifique. On pourrait alors avoir le sentiment que le mérite est une idée consensuelle et omniprésente parce qu'il s'agit d'une idée complètement vide. Puisqu'il est possible de justifier n'importe quoi au nom du mérite, il ne peut s'agir d'une idée fondamentale, puisque, à première vue, une idée fondamentale doit posséder un contenu spécifique. Les idées d'égalité ou de justice possèdent un certain contenu. On pourrait craindre que le mérite soit une idée trop vaste et trop large pour en posséder un.

Cependant, de tels doutes sont probablement exagérés. Même si le terme de mérite peut être utilisé dans tous les sens dans la langue courante, cela ne veut pas dire qu'il ne soit pas possible de trouver un sens plus limité, correspondant à une idée morale précise¹. De fait, l'idée qu'un individu mérite un bon salaire semble posséder un contenu conceptuel fort. Dire qu'il a besoin de ce salaire, ou qu'il serait bon qu'il le reçoive, renvoie à d'autres idées. S'il le mérite, c'est qu'il a fait quelque chose qui le met en droit de le recevoir, probablement pour une raison précise. On semble quitter le caractère vague associé à l'idée que tout être humain mérite d'être heureux.

Pourtant, l'idée qu'un individu mérite un salaire semble également receler de nombreuses difficultés. La première difficulté, bien connue, tient aux multiples raisons pour lesquelles on peut dire qu'un individu mérite un salaire. Est-ce que cette personne mérite ce salaire pour ses efforts, pour sa contribution, pour compenser le fait que le travail est potentiellement pénible ou peu valorisant ? Cette pluralité des bases du mérite n'est pas un problème en soi ; cependant, elle semble problématique en ce qu'aucune réponse possible ne semble parfaitement satisfaisante.

Dire qu'un individu doit être récompensé uniquement pour ses efforts peut apparaître plausible à première vue ; cependant, de nombreux exemples montrent bien qu'un tel critère peut mener à des situations extrêmement peu intuitives. Imaginons que deux employés font exactement les mêmes efforts, mais que l'un des deux est extrêmement maladroit, causant de nombreux dommages pour l'entreprise, tandis que l'autre est extrêmement habile et compétent, conduisant à une production accrue et de qualité. Il semble assez peu vraisemblable de dire que les deux méritent véritablement d'être payés le même salaire, ou qu'ils méritent un salaire identique.

¹ Le chapitre 1 sera justement consacré à la délimitation d'un sens précis du mérite.

Le critère de la contribution pourrait alors être mis en avant, mais il laisse rapidement place à de nombreuses difficultés. La plus importante tient à la difficulté de mesurer la contribution, à trouver un critère objectif qui permette d'identifier exactement ce que doit recevoir chaque individu. En un mot, les théories de la contribution se tiennent souvent face à un dilemme : on peut formuler un critère objectif, mais dans ce cas ce critère est souvent extrêmement problématique ; ou bien on peut laisser l'idée de contribution être évaluée de façon floue et indéterminée. Sur la première branche du dilemme, un critère très objectif tient à l'idée que la contribution des individus est identifiable aux salaires qu'ils sont capables de recevoir dans un marché libre et non régulé. Mais ce critère semble très éloigné de nos jugements ordinaires : le salaire qu'un individu doit obtenir sur un marché dépend de nombreux facteurs extérieurs, comme l'offre et la demande, qui feraient que ce que mérite un individu relèverait très largement de la chance. Sur un marché libre, un fermier peut se voir obtenir un salaire très faible simplement parce que le prix des denrées a chuté. L'idée que le mérite puisse être assimilé à des facteurs aussi volatiles semble très improbable². D'un autre côté, si l'on renonce au critère du marché pour identifier ce que mérite chaque individu, on se retrouve face à des critères très flous, tels que l'idée qu'un individu a fait un « bon » travail, avec de grosses difficultés : comment comparer deux occupations très différentes, par exemple dans des secteurs différents de l'économie ? Comment comparer le travail d'un ouvrier, d'un commerçant, d'un agriculteur ? De telles difficultés ne sont peut-être pas insurmontables mais, au minimum, complexifient fortement la question.

Enfin, toujours en ce qui concerne l'identification des critères du mérite, on peut également noter que la notion a historiquement été très fortement associée à un tout autre type de critères, à savoir la vertu ou l'excellence morale d'un individu. Le mérite a fait l'objet de débats théologiques importants liés à la question de savoir quels individus méritaient le paradis, ou, même dans un contexte laïque, liés à la question de déterminer dans quelle mesure seuls des individus vertueux méritent d'être heureux³. Savoir dans quelle mesure l'excellence morale doit être prise en compte dans le mérite complexifie donc encore la question. Dans un contexte

² Cet argument a été mis en avant avec force par Sidgwick ; Hayek propose également une très bonne version de l'argument. Henry SIDGWICK, *The Methods of Ethics*, Indianapolis, Hackett Publishing Company, 1981 ; F. A. HAYEK, *Law, Legislation and Liberty*, London, Routledge, 2012.

³ Pour les débats théologiques liés à la relation entre mérite et souverain bien, la discussion de Thomas d'Aquin dans la Somme théologique reste extrêmement intéressante. Sur la question du mérite et du bonheur, les discussions de Kant et de Ross continuent d'être citées dans les débats contemporains. THOMAS, *Textes sur la morale*, Paris, J. Vrin, 2011 ; Immanuel KANT, *Fondements de la métaphysique des mœurs*, trad. fr. Victor DELBOS, Paris, Librairie Générale Française, 1993 ; David ROSS, *The Right and the Good*, Oxford, Clarendon Press, 2003.

contemporain, on peut penser à la question de savoir dans quelle mesure il serait important de récompenser les étudiants méritants.

Mais même si la question du critère du mérite avait été résolue, un autre problème se pose, lié à la détermination des critères de responsabilité nécessaires pour que le critère de mérite apparaisse comme un critère juste. Il semble apparemment intuitif qu'un individu doit pouvoir être responsable de ses actes pour qu'il puisse mériter quelque chose. Je ne peux rien mériter si je n'ai pas été responsable des raisons qui me font méritant. Ce lien entre mérite et responsabilité pose problème, du fait de deux difficultés, l'une métaphysique, l'autre pratique.

La question métaphysique tient bien sûr à la question de savoir dans quelle mesure des individus peuvent être considérés comme libres et responsables dans un univers dont les lois causales les dépassent largement. Il s'agit bien sûr de la question traditionnellement abordée sous l'angle du lien entre déterminisme et existence du libre-arbitre. Même si résoudre ce problème entraîne le lecteur très loin des questions traditionnelles dans la réflexion politique, il est évident qu'une solution satisfaisante des questions liées au mérite implique une résolution de la question métaphysique⁴.

La question pratique tient à savoir dans quelle mesure il est juste de récompenser des individus pour des activités méritoires, sachant qu'il est probable que la chance, ou le hasard, sont intervenus dans une grande partie pour rendre possibles ces activités méritoires. Plus précisément, cette question se pose de façon particulièrement accrue ce qui concerne la rupture de l'égalité existant entre deux individus. Imaginons qu'un employé, disons Paul, a effectué un très bon travail. Mérite-t-il d'obtenir un très bon salaire de ce fait ? Si l'on prend la question de façon non-comparative, la réponse semble clairement positive. Mais le cas devient beaucoup plus délicat si l'on prend en compte toutes les autres personnes qui auraient pu effectuer ce travail, mais qui ne l'ont pas fait pour des raisons de hasard et de pures circonstances. Imaginons qu'une autre personne travaillant dans le même secteur que Paul est au chômage pour des raisons totalement indépendantes de sa volonté. Cette deuxième personne, que l'on peut appeler Pierre, aurait pu parfaitement effectuer le même travail que Paul, si jamais elle avait été placée dans les mêmes circonstances. Dans cette situation, peut-on dire que Paul mérite de gagner davantage d'argent que Pierre, alors même que seul le hasard est responsable de leur différence de situation ? Le lecteur pourra être tenté de répondre par la négative. Mais dans ce cas, si l'on affirme que les deux individus méritent de gagner le même salaire, il devient presque impossible

⁴ Pour un exemple d'un auteur défendant l'idée que les théories liées au mérite et à la responsabilité sont inadéquates parce qu'incompatibles avec des conceptions plausibles du libre-arbitre, voir Jean-Fabien SPITZ, *Abolir le hasard? Responsabilité individuelle et justice sociale*, Paris, Librairie Philosophique Vrin, 2008.

de déterminer ce qu'un individu mérite réellement de gagner au sein de la société. Le hasard semble s'être mêlé à trop de faits, dans la détermination génétique, dans l'influence familiale, dans les rencontres faites par chaque personne, pour que l'on puisse dire avec certitude ce qu'un individu mérite. Le mérite ne deviendrait pas alors tant impossible que proprement inapplicable⁵.

Tous ces doutes pourraient sans doute être remis en question s'il était possible de fournir un fondement normatif fort à l'idée de mérite. Mais expliquer pourquoi il est important de récompenser les individus selon leur mérite est en soi un problème difficile à résoudre. Il est bien sûr possible de fournir une réponse simple à cette question : récompenser les individus selon leur mérite serait une manière de les inciter à agir de façon productive ou adéquate à l'avenir. La récompense au mérite serait alors une manière de modifier le comportement ultérieur des individus afin de servir des objectifs d'utilité sociale, comme la production ou l'efficacité⁶. Une telle justification du mérite peut bien sûr servir à expliquer ou à assurer la défense d'un certain nombre de nos pratiques quotidiennes. Mais cette justification revient alors à dénier toute valeur propre au mérite, pour la reléguer dans un rôle purement instrumental. La véritable valeur serait alors dans une autre idée morale, comme l'égalité, le soutien au plus défavorisés, ou le bien-être collectif, et les pratiques de récompense au mérite ne seraient dans ce cas que des moyens pour atteindre ces fins. Dans ce cas, si l'on n'accorde au mérite qu'une fonction instrumentale, on pourrait alors considérer que la récompense au mérite ne serait pas vraiment juste, en tant que d'autres systèmes pourraient peut-être contribuer à ces objectifs d'efficacité ou de soutien aux plus pauvres sans contribuer à créer des inégalités, comme la récompense au mérite semble le faire⁷. Justifier le mérite comme un instrument semble, en fin de compte, une manière de délégitimer le mérite comme principe moral important.

A l'inverse de ces justifications purement instrumentales des récompenses au mérite, on peut bien sûr considérer que le mérite est intuitivement une idée morale forte. Mais il n'est pas évident qu'un pur appel aux intuitions suffise à justifier la notion. En effet, l'on peut considérer que l'intuitivité d'une idée morale ne la rend valide que si cette idée morale ne se heurte pas par ailleurs à d'autres contre-arguments ou difficultés. Or, comme nous l'avons vu

⁵ C'est bien sûr le cœur de la critique célèbre effectué par Rawls. John RAWLS, *Théorie de la justice*, Paris, Points, 2009.

⁶ Une plus grande efficacité qui peut ensuite servir d'autres buts, comme dans la théorie rawlsienne, où la plus grande efficacité due à des incitations méritocratiques sert *in fine* à améliorer le sort des plus défavorisés.

⁷ On peut penser aux arguments défendus par G. A. Cohen, qui mettait en avant qu'une société véritablement juste serait une société capable de créer éthos égalitaires, où les individus seraient prêts à travailler pour le bien collectif sans exiger en contrepartie des salaires extravagants. G. A. COHEN, *If You're an Egalitarian, How Come You're so Rich?*, Harvard University Press, 2001.

précédemment, le mérite semble assailli de différents problèmes. De plus, on peut aussi considérer qu'une idée morale intuitive ne possède véritablement de la valeur que si elle est compatible avec les autres idées morales auxquelles nous tenons. Dans le cas du mérite, on peut considérer que l'idée entre en conflit avec l'égalité, et doit être mise de côté s'il n'est pas possible de trouver de justifications indépendantes à l'idée de mérite. Or, de nombreux philosophes ont argumenté en faveur de l'idée que les deux seuls arguments en faveur du mérite tiennent à son intuitivité et à son rôle incitatif. De telles justifications peuvent sembler faibles au regard des obscurités liées à la notion.

Récapitulons maintenant les différents arguments mobilisés à l'encontre de l'idée de mérite. Tout d'abord, comme nous l'avons vu, le mérite souffre de son caractère terriblement vague. L'idée semble pouvoir être mobilisée pour justifier n'importe quelle proposition. Deuxièmement, il est difficile de déterminer quel devrait être le critère du mérite. La notion semble également indéterminée de ce point de vue. Troisièmement, le fait qu'un jugement de mérite doit reposer sur la responsabilité d'un individu pose deux difficultés, l'une métaphysique, liée à la relation entre libre-arbitre et déterminisme, l'autre pratique, liée à la difficulté d'estimer le mérite des individus dans un contexte où la chance et le hasard ont presque à coup sûr eu un impact décisif sur la vie des individus. Quatrièmement, il n'est pas évident de fournir une justification du mérite qui ne soit pas ou bien purement intuitive ou bien purement instrumentale, et il n'est pas clair que ces deux types de justifications puissent constituer une défense véritable du mérite en tant que tel. Enfin, cinquièmement, le mérite semble entrer en conflit avec d'autres idéaux auxquels nous tenons, comme l'idée d'égalité.

Toutes ces difficultés ont été énoncées de façon limpide dans la littérature philosophique de langue anglaise, et ont connu comme moments forts les critiques portées par Henry Sidgwick en 1874 dans *Les méthodes de l'éthique*, et les critiques puissantes portées par Friedrich Hayek et John Rawls dans les années 1960-1970. De fait, la plus grande partie des théories influentes à la fin du XXe siècle ont rejeté l'idée comme incohérente, impraticable, ou inacceptable d'un point de vue moral. A partir des années 60, les plus grandes théories de la justice ont tour à tour rejeté la valeur de cette notion, aboutissant à un quasi consensus sur l'absence d'importance du mérite. De 1965 à 1980, Brian Barry, John Rawls, Ronald Dworkin, Thomas Nagel, Robert Nozick, ont construit des théories de la justice n'accordant aucune importance à la notion⁸. La

⁸ Scheffler a bien montré qu'aucune des théories contemporaines dominantes de la justice n'accorde une place importante au mérite. Dans le paysage anglo-saxon, seule la théorie de David Miller fait en effet exception, en soutenant une défense élaborée du mérite. Le paysage français n'est pas très différent : parmi les publications récentes, Spitz, Michaud et Savidan, à partir de prémisses différentes, rejettent tous le mérite comme un concept dépourvu de validité. Ce rejet au sein des théories de la justice sociale contraste avec un soutien plus ferme dans

situation a depuis peu changé, même si ces dernières années ont vu un certain renouveau de théories défendant le mérite, notamment dans le cadre du succès des théories de l'égalitarisme de la fortune⁹.

Il existe donc une panoplie impressionnante de critiques possibles du mérite, qui amènent à douter du fait que la notion ait même un sens. Ces critiques ont eu un succès remarquable au cours des dernières décennies, aussi bien dans les milieux philosophiques que dans d'autres domaines des sciences sociales. J'argumenterai dans cette thèse, cependant, que ces doutes relatifs au mérite sont en réalité très nettement exagérés, et qu'il est tout simplement impossible de se passer de cette notion.

Construire une défense du mérite à partir du sens commun

Chacune des critiques envisagées précédemment fera l'objet d'une réponse spécifique. Dans ce cadre, je présenterai une défense générale de l'idée de mérite. Cette défense passera d'abord par une analyse conceptuelle qui permettra d'aboutir à un concept spécifique de mérite, ce qui permettra de réfuter les critiques reprochant le caractère irrémédiablement vague de l'idée de mérite. Cette analyse conceptuelle du mérite nous occupera dans la première partie de cette thèse (chapitres 1 à 3). Dans une deuxième partie, il s'agira de clarifier la méthodologie philosophique employée ici, à partir d'un examen des raisons pour lesquelles une philosophie normative doit se fonder sur la considération des opinions courantes (chapitres 4 et 5). La troisième partie constituera un examen des fondements de l'idée de mérite, afin de produire une justification qui se basera sur quatre raisons principales (chapitres 6 à 9). Cette justification du mérite visera à répondre aux critiques reprochant à l'idée son faible fondement normatif. Dans

d'autres sections de la philosophie ou au sein d'autres disciplines. Samuel SCHEFFLER, « Responsibility, Reactive Attitudes, and Liberalism in Philosophy and Politics », *Philosophy and Public Affairs*, 1992, n° 21, p. 299-323 ; J.-F. SPITZ, *Abolir le hasard?*, *op. cit.* ; Yves MICHAUD, *Qu'est-ce que le mérite ?*, Paris, Folio, 2011 ; Patrick SAVIDAN, *Repenser l'égalité des chances*, Paris, Hachette Littératures, 2010.

⁹ Parmi les égalitaristes de la fortune ayant cherché à distinguer une place importante au mérite, on peut compter notamment Arneson et Temkin. En dehors de David Miller, qui a sans doute été l'auteur le plus important dans l'articulation d'une défense du mérite, on peut compter David Schmidtz. Il reste cependant correct de dire que le mérite reste une notion rejetée par la majorité des grands théoriciens contemporains dans le domaine de la justice sociale. Richard ARNESON, « Luck Egalitarianism—A Primer », in Carl KNIGHT et Zofia STEMPOWSKA (dir.), *Responsibility and Distributive Justice*, Oxford University Press, 2011, p. 24–50 ; Larry TEMKIN, « Justice, Equality, Fairness, Desert, Rights, Free Will, Responsibility, and Luck », in Carl KNIGHT et Zofia STEMPOWSKA (dir.), *Responsibility and Distributive Justice*, Oxford University Press, 2011, p. ; David MILLER, *Principles of Social Justice*, Harvard University Press, 2001 ; David SCHMIDTZ, *The Elements of Justice*, Cambridge, Cambridge University Press, 2006.

une quatrième partie (chapitres 10 à 12), je défendrai l'idée que le mérite repose en réalité sur une conception simple et consensuelle de l'idée de responsabilité. Enfin, dans la conclusion, il s'agira de défendre l'idée d'une compatibilité entre mérite et égalité.

Cette articulation générale d'une théorie du mérite se fera bien sûr progressivement au cours de cette thèse. Il s'agit cependant ici de mettre en avant un aspect clé de la méthode employée, méthode qui devrait permettre de montrer pourquoi les critiques précédemment mentionnées du mérite sont erronées, ou du moins peu plausibles. Dans cette thèse, nous défendrons l'idée que la philosophie politique doit se fonder sur une connaissance approfondie des idées morales ordinaires, telles que ces idées peuvent être reconstituées à partir d'enquêtes psychologiques ou sociologiques. Comme John Rawls le défendait il y a quarante ans, nous considérons en effet que toute théorie politique normative doit consister avant tout en une rationalisation et une épuration des idées morales ordinaires¹⁰. Si l'on adopte cette perspective, il devient difficile de défendre la possibilité de construire sans avoir pris en compte les idées courantes, c'est-à-dire la manière dont les membres de nos sociétés considèrent l'idée de mérite. Et selon nous, prendre en compte ces idées ordinaires conduit à la conclusion que le mérite est une idée indispensable et parfaitement valable, contre laquelle il n'existe fondamentalement aucun argument plausible.

En défendant l'importance des idées courantes pour l'activité philosophique, je m'inscris dans le champ en plein développement de la philosophie expérimentale. Né au début des années 2000 dans les pays anglo-saxons, et maintenant présent dans de nombreux pays, ce champ d'étude pluridisciplinaire consiste à mobiliser des méthodes empiriques pour étudier les idées courantes relatives aux problèmes philosophiques classiques¹¹. De fait, les méthodes employées en philosophie expérimentale ont souvent consisté à mobiliser les méthodes utilisées en psychologie sociale, notamment l'usage de vignettes et d'histoires visant à étudier les jugements relatifs à des situations imaginaires. Depuis vingt ans, la philosophie expérimentale a enquêté sur les idées relatives à de nombreux champs classiques, avec un intérêt particulier porté à la philosophie de l'action, à la philosophie morale, à l'épistémologie, et aux théories du libre-arbitre.

Dans le cadre de la philosophie normative, je défendrai dans les chapitres 4 et 5 de cette thèse l'idée que toute théorie doit partir des jugements les plus courants, en tant que ceux-ci

¹⁰ John RAWLS, « The Independence of Moral Theory », *Proceedings and Addresses of the American Philosophical Association*, 1974, vol. 48, p. 5-22.

¹¹ Pour une introduction à la philosophie expérimentale, on peut consulter en Français l'anthologie des articles les plus influents des années 2000, et un livre introductif écrit par Florian Cova. Florian COVA (dir.), *La philosophie expérimentale*, Vuibert, 2012 ; Florian COVA, *Qu'en pensez-vous ? Introduction à la philosophie expérimentale*, Meaux, Germina, 2011.

constituent le fondement de toute vie morale partagée. Dans le cas du mérite, cela signifie que les idées de mérite possèdent une forte plausibilité à première vue, étant donné qu'elles font l'objet d'un très fort consensus, aussi bien en France que dans un très grand nombre de pays. Cependant, cela ne veut pas dire que les philosophes sont condamnés à adopter les idées de la plupart des gens et à chercher à justifier ces idées telles quelles. Nous comptons rester fidèles dans notre thèse à la méthode désormais classique que Rawls, dans la *Théorie de la justice*, a appelé l'équilibre réflexif. L'équilibre réflexif correspond au travail philosophique de mise en adéquation entre, d'un côté, une théorie morale, et, de l'autre côté, les intuitions éthiques ordinaires, conduisant un individu à faire un aller-retour entre ces deux niveaux. Les intuitions morales doivent être prises en compte, parce qu'elles constituent la substance et la base des idées morales d'un individu ; la théorie morale, cependant, possède également sa propre valeur, parce qu'elle permet de mettre en cohérence entre elles les différentes intuitions morales, afin d'éliminer celles qui, après réflexion, se révéleraient être de purs préjugés. Il n'est évidemment pas sûr que toutes les intuitions doivent être considérées comme valables ; d'un point de vue normatif, nous serons potentiellement contraints de réviser nos idées à propos du mérite. Cependant, il est impossible de préconiser cette révision avant même de savoir quelle est l'opinion du sens commun en la matière. Et, dans le cas particulier du mérite, nous défendrons l'idée que les opinions du sens commun sont très largement correctes, et accordent une place centrale au mérite dans notre univers moral.

Première partie

Fondations conceptuelles de l'idée de mérite

Chapitre 1

Définir le mérite

Les usages ordinaires du mérite présentent une très grande diversité. Dès que l'on prête attention aux usages multiples du mot « mérite » et des expressions associées, on ne peut qu'être frappé par l'omniprésence de ces termes, sans qu'il soit évident qu'une signification bien spécifique y soit toujours associée. On parle de « pauvre méritant », « d'étudiant méritant », de mériter une peine, une récompense, une chance. A travers la multiplicité des usages, il est ardu de voir en quoi le mérite peut prendre une consistance forte qui le transforme en un terme pertinent pour décrire les pratiques morales quotidiennes. Ce chapitre a pour but de procéder à un examen des grandes caractérisations philosophiques du mérite, afin de sélectionner les traits pertinents du concept.

Un tel examen de caractérisation et de définition du mérite est rendu d'autant plus nécessaire qu'il existe un foisonnement de théories philosophiques à son sujet, sans qu'il soit toujours bien évident que différents philosophes parlent de la même chose. Le mérite a été tour à tour assimilé à l'intégralité de la justice, à la compensation des désavantages subis par une personne, à la récompense d'une vertu supérieure, ou à la gratitude exprimée vis-à-vis d'actions passées. Il s'agira ici de montrer que toutes ces caractérisations possibles du mérite ne sont pas aussi pertinentes d'un point de vue philosophique. Il nous faut construire un concept du mérite qui possède d'une part une certaine correspondance avec les usages quotidiens du mot, et d'autre part un contour précis et déterminé permettant un traitement philosophique approprié. Il s'agira donc de maintenir un équilibre entre la fidélité aux usages quotidiens, malgré leur caractère flou parfois problématique, et la rigueur d'un traitement philosophique, qui risque d'être stérile s'il n'a aucun accord avec les pratiques et les idées quotidiennes habituellement associées à l'idée de mérite.

Ce chapitre procédera en trois temps. Dans un premier temps, je montrerai la diversité radicale des emplois des termes « mérite » et « mériter » dans le langage ordinaire, et j'indiquerai pourquoi je considère qu'il est important de restreindre la diversité des usages.

Dans un deuxième temps, à partir notamment des théories développées par Joel Feinberg, David Miller, et Serena Olsaretti, je proposerai une caractérisation minimale du mérite, et j'indiquerai les conditions indispensables que doit satisfaire toute théorie du mérite, si elle veut à la fois être plausible, conforme aux usages quotidiens, et spécifique. Dans un troisième temps, je montrerai que cette caractérisation du mérite permet déjà d'exclure tout un ensemble de théories philosophiques du mérite comme étant non-pertinentes, parce que ne satisfaisant pas des critères minimaux de plausibilité ou de spécificité.

La diversité des usages des termes « mérite » et « mériter »

Pour voir la très grande diversité des usages possibles du terme de mérite, nous pouvons par exemple prendre la liste dressée par George Sher au début de son livre *Desert* :

- « 1. Jones mérite sa réussite ; il a travaillé dur.
2. Smith méritait plus de succès qu'il n'en a obtenu ; il a donné tout ce qu'il avait.
3. Walter mérite le poste ; c'est le candidat le plus qualifié.
4. Wilson méritait d'être disqualifié ; il savait que la date limite pour les candidatures était le premier mars.
5. Jackson mérite plus que le salaire minimum ; son travail est important et il le fait bien.
6. Baker mérite de gagner ; il a très bien joué.
7. Miss Vermont mérite de gagner ; c'est la plus jolie candidate.
8. Anderson mérite sa condamnation à vingt-deux ans de prison ; il a planifié l'assassinat.
9. Brown savait peut-être qu'il ne serait pas pris, mais il mérite quand même d'être puni.
10. Winters mérite d'être dédommagé ; il a souffert de douleurs constantes depuis la fusillade.
11. Lee mérite une récompense ; il a risqué sa vie.
12. Benson mérite d'avoir plus de chance à l'avenir ; c'est quelqu'un de bien.
13. Gordon mérite d'avoir plus de chance à l'avenir ; il n'a pas eu de veine jusqu'à présent.
14. McArthur mérite d'être interrogé ; c'est un expert sur le sujet.
15. Cleveland mérite une meilleure publicité ; c'est une ville intéressante¹². »

¹² George SHER, *Desert*, Princeton, NJ, Princeton University Press, 1989, p. 7.

Ces quelques phrases montrent bien la diversité radicale des usages du mot « mériter ». Tous ces exemples possèdent certes une certaine unité conceptuelle : chaque phrase traduit l'idée qu'il serait bon qu'un individu, ou qu'une chose, reçoive quelque chose ayant une valeur positive ou négative. Il s'agit bien d'une signification partagée, mais il s'agit d'un squelette conceptuel ayant une très faible valeur. Le mot mérite correspond ici à l'idée qu'il est approprié (« fitting » dans la littérature anglo-saxonne) que quelqu'un reçoive un bien ou un mal. Cette idée d'appropriation n'est pas tout à fait vide. Cependant, dans les exemples cités ci-dessus, les raisons sous-jacentes à ce caractère « approprié » possèdent une diversité très forte.

On court le risque, souligné par Brian Barry dans son livre *Political Argument*, de ne voir dans l'idée de mérite qu'une notion absolument vide, qui traduise uniquement le fait qu'il serait bon que quelqu'un obtienne quelque chose, sans qu'un principe de justice strict soit associé à l'idée de mérite elle-même¹³. Le mérite se confondrait avec l'ensemble de la justice, voire avec l'ensemble de la moralité. Il est crucial de restreindre le concept de mérite, pour voir en quoi celui-ci possède une certaine spécificité, et n'est pas simplement identique à l'idée que quelque chose soit bon, ou approprié. Ce chapitre sera donc consacré à la détermination d'un concept précis de mérite, qui possède une certaine substance, correspondant à ses usages politiques actuels.

Dans cette détermination d'un sens plus précis du mérite, il ne s'agit pas de dire que le langage ordinaire est systématiquement erroné, ou qu'il existe un sens unique, philosophique, du mérite qui serait le bon. Il s'agit plutôt de dire que, même s'il existe une multiplicité de sens du mérite dans le langage ordinaire, une bonne partie n'a pas grand intérêt d'un point de vue philosophique, et ne capture pas ce qui est véritablement intéressant et problématique avec l'idée de mérite. Je reprends ici la conclusion de Serena Olsaretti dans son livre *Liberty, Desert and the Market* : « Je ne pense pas que nous devrions accepter que tous ces usages du mérite sont également pertinents pour la justice, et nous ne devrions pas essayer de montrer que certains usages en apparence sensés du mérite n'ont pas de sens. Je suggère plutôt que nous

¹³ Cette crainte est évoquée dans la nouvelle préface de la deuxième édition de Brian BARRY, *Political Argument: A Reissue With a New Introduction*, Berkeley, Univ of California Press, 1990. Une théorie philosophique récente qui identifie le mérite à l'idée qu'il est bon de récompenser quelqu'un, quelque soit la raison pour cela, est offerte par Fred FELDMAN, *Distributive Justice: Getting What We Deserve From Our Country*, Oxford, OUP Oxford, 2016. Même s'il identifie un mérite politico-économique qui est censé constituer une forme précise du mérite, sa perspective reste extrêmement large et vague : « Au plus haut niveau d'abstraction, nous trouvons le même couple de facteurs [dans tous les jugements de mérite politico-économique] : être un être humain ordinaire qui possède des besoins essentiels d'un point de vue de la communauté en tant que fondement du mérite, et être ancré dans une communauté qui essaiera de s'assurer que ces besoins sont satisfaits en tant que chose méritée. » ; *Ibid.*, p. 71. Au-delà des termes spécifiques employés par Feldman, ce dernier tend à caractériser comme besoins un ensemble extrêmement large de biens que la communauté peut prodiguer à l'individu. On voit là une assimilation du mérite à la justice toute entière, au point que la notion perd toute spécificité.

délimitations progressivement la notion de mérite qui nous intéresse, en identifiant certains desiderata ou contraintes que le principe de mérite doit satisfaire afin d'être défendable¹⁴. »

Prenons quelques exemples pour illustrer la diversité d'usages et de connotations. Dans les différentes phrases tirées du livre de George Sher, on trouve presque toujours l'idée que quelqu'un doit recevoir une bonne chose. Par contre, les raisons pour lesquelles les individus doivent recevoir une bonne chose divergent. Dans un cas, par exemple, « Gordon mérite d'avoir plus de chance à l'avenir ; il n'a pas eu de veine jusqu'à présent. », la raison pour laquelle Gordon devrait recevoir un bien réside dans le fait qu'il a subi quelque chose ; dans le cas de « Walter mérite le poste ; c'est le candidat le plus qualifié. », la raison pour laquelle il serait bon que Walter obtienne le poste tient à un trait personnel de Walter, le fait qu'il possède une certaine qualité. Enfin, dans le cas de « Baker mérite de gagner ; il a très bien joué. », la raison du jugement tient dans le fait que Baker a effectué une action positive. Dans les trois cas, nous avons donc un jugement de mérite qui conclut au fait qu'un individu devrait recevoir un bien ; cependant, à chaque fois, la raison sous-jacente au fait qu'il doive recevoir un bien correspond à une raison différente. Dans le premier cas, le cas de la malchance, la raison du mérite tient à une histoire passée, pour laquelle l'individu n'est pas responsable. Le fait qu'un individu ait subi de la malchance n'indique rien de particulier sur cette personne, et ne traduit aucun caractère positif ou négatif associé à l'individu. Dans le second cas, celui de l'individu méritant un emploi du fait de ses qualifications, le mérite tient au fait qu'un individu possède des qualités propres. Dans le troisième cas, celui du sportif, le mérite tient, non plus à une qualité inhérente à l'individu, mais à une action passée. Les trois raisons du mérite sont donc l'histoire individuelle, les qualités propres à une personne, et les actions passées. Il est, là encore, difficile de voir en quoi le mérite correspond à une idée précise, spécifique.

Comment découvrir ce qui est propre, spécifique à l'idée de mérite ? Comment obtenir un concept qui ne se réduise pas à l'idée de justice ? Il me semble que l'usage courant du terme de mérite présente au moins trois caractéristiques qui le distinguent en général du terme de justice. Tout d'abord, le mérite renvoie à l'excellence d'une action ou d'un caractère ; ou plutôt, à défaut d'excellence, au caractère en tout cas positivement valorisé. Cette première caractéristique en implique une seconde : puisque le mérite se fonde d'abord sur une excellence, il n'est pas donné de premier abord que toutes les actions, ou tous les caractères, possèdent cette excellence au même degré. Il serait même étonnant que cela soit le cas. En conséquence, le mérite offre toujours la possibilité d'une inégalité entre les individus. Enfin, le mérite

¹⁴ Serena OLSARETTI, *Liberty, desert and the market: a philosophical study*, Cambridge, Cambridge Univ. Press, 2004, p. 11.

correspond non seulement à l'idée qu'il existe des différences de qualité entre les actions ou les caractères, mais implique de plus qu'il serait bon de récompenser de façon inégale les individus concernés, et que chacun obtienne une récompense en proportion de la qualité de ses actions ou de son caractère¹⁵.

A partir de ces trois traits caractéristiques, on comprend pourquoi le concept de mérite est, fondamentalement, un concept controversé, et que les connotations associées ne sont pas celles d'un concept correspondant à l'idée de justice ou de moralité. Dans la sphère politique, comme dans la plupart des débats philosophiques à son sujet, les connotations associées au mérite renvoient à l'idée qu'il serait possible de justifier des inégalités du fait de la contribution supérieure de certains membres de la société. Récompenser les individus selon leur mérite correspondrait alors au fait qu'il est bon que les individus obtiennent un bien correspondant à leurs actions passées. Un concept opérationnel de mérite doit donc posséder les conséquences suivantes : premièrement, il doit être fondé sur des actions, positives ou négatives, passées d'un individu ; deuxièmement, il doit pouvoir justifier des inégalités. Indépendamment de savoir si le concept de mérite possède ou non une valeur, une conceptualisation du mérite qui ne satisferait pas ces deux réquisits serait vraisemblablement trop éloignée des usages actuels du terme « mérite » pour valoir la peine d'être débattue dans un contexte philosophique ou politique.

Pour affiner notre caractérisation précise du mérite, il est utile de revenir sur certaines grandes conceptualisations du concept qui ont été effectuées dans la littérature philosophique. Comme tant d'autres sujets centraux dans la philosophie politique et morale, une étude précise du concept de mérite a connu un essor important à partir du renouveau de la philosophie politique et morale anglo-saxonne à partir des années 1960-1970. En ce qui concerne la question du mérite, deux travaux ont été particulièrement importants dans ce renouveau, et sont indispensables pour développer une caractérisation plus fine de l'idée de mérite : l'article « Justice and Personal Desert » de Joel Feinberg, publié en 1967, et naturellement les chapitres

¹⁵ Cette idée d'une récompense proportionnelle à la qualité des actions des individus est une idée assez ambiguë. D'un point de vue mathématique, l'idée de proportion renvoie à l'idée qu'il est possible de trouver une correspondance exacte entre la qualité des actions et la qualité des récompenses, et qu'il est possible de définir ce que cela veut dire que d'être deux fois plus méritant que quelqu'un d'autre. Je reviendrai plus tard sur cette question. Pour l'instant, le lecteur peut simplement interpréter l'idée de proportionnalité entre mérite de l'action et récompense comme une certaine idée d'ordre : un plus grand mérite implique une plus grande récompense, même s'il n'est pas possible de déterminer avec une précision mathématique en quoi la récompense est « proportionnelle » au mérite.

que Rawls consacre à l'idée de mérite dans *Théorie de la justice* en 1971¹⁶. Je reprendrai à tour de rôle l'étude de ces deux textes, et les réponses qu'ils ont inspirées.

Une première conceptualisation du mérite

Une première manière d'obtenir une caractérisation plus précise de l'idée de mérite consiste à passer par un début de formalisation, ou d'abstraction. Dans un article célèbre, repris dans son livre *Justice and Personal Desert*, Joel Feinberg caractérise le fait de mériter quelque chose comme un état à la structure triadique : A mérite B en vertu de C, où A doit être un agent, B un bien ou un mal, et C un fait caractérisant l'individu en question. A est alors le sujet du mérite, B l'objet du mérite, et C la base ou le fondement du mérite¹⁷.

Une telle caractérisation reste extrêmement large. Nous proposerons deux amendements nécessaires. Comme nous l'avons vu plus haut, je considère que le mérite renvoie généralement à l'idée d'une certaine qualité positive, ou d'une excellence. En conséquence, si l'on reprend la structure proposée par Feinberg, il faut dire que C, la raison du mérite, renvoie nécessairement à un fait caractérisant l'individu qui dénote un certain degré de qualité ou d'excellence. Autrement dit, la raison du mérite doit toujours renvoyer à un fait positif caractérisant l'individu.

Même si la structure triadique proposée par Feinberg a été très largement reprise, elle a été récemment critiquée par Fred Feldman, notamment dans son livre *Distributive Justice*¹⁸. Dans ce livre, Feldman montre qu'il n'y a pas de raison de se limiter à une structure triadique comme celle proposée par Feinberg. Feldman propose de considérer l'ensemble des caractéristiques possibles d'un jugement de mérite, qui n'impliquent pas moins de six éléments différents : 1) une personne méritante, 2) une chose méritée, 3) une personne qui doit distribuer la chose méritée, 4) la force du mérite, 5) le temps et les circonstances de distribution, et 6) la raison du mérite. Par exemple, il est possible de dire que Jean (personne méritante) mérite que Pierre (distributeur) lui apporte de l'aide dans son déménagement (chose méritée) jeudi à Paris (temps et circonstances), parce Jean avait aidé Pierre lors d'une précédente occasion (raison). On peut

¹⁶ Joel FEINBERG, *Doing and Deserving: Essays in the Theory of Responsibility*, Princeton, Princeton University Press, 1974 ; J. RAWLS, *Théorie de la justice*, op. cit.

¹⁷ J. FEINBERG, *Doing and Deserving*, op. cit.

¹⁸ F. FELDMAN, *Distributive Justice*, op. cit.

rajouter une information ajoutant qu'il serait fortement injuste que Pierre n'aide pas Jean (force du mérite).

Toutes ces caractérisations sont intéressantes, et permettent d'obtenir une vue synthétique de différents éléments présents dans une bonne partie des jugements de mérite. Cependant, on peut considérer que de telles caractérisations formelles ne nous aident pas vraiment à voir ce qu'il y a de spécifique dans les jugements de mérite. La plupart des jugements moraux possèdent, par exemple, un sujet, un objet, une raison, un temps et des circonstances. Pour obtenir une caractérisation plus précise des jugements de mérite, il faut ajouter des critères plus spécifiques au mérite à proprement parler.

Dans ce cadre, je propose de reprendre une partie des critères adoptés par Serena Olsaretti dans son livre *Liberty, Desert and the Market*, déjà cité. Les premières caractéristiques défendues par Olsaretti font généralement consensus, et correspondent à des idées que nous avons déjà exprimées. Tout d'abord, le mérite doit traduire des caractéristiques propres à une personne, qu'il s'agisse de traits caractériels, de capacités, ou d'actions effectuées par un sujet. Ensuite, le mérite doit posséder un caractère évaluatif : nous ne disons pas qu'un individu a du mérite si ses actions sont parfaitement neutres ; celles-ci doivent être ou bien positives, ou bien négatives. Enfin, selon Olsaretti, les jugements de mérite doivent posséder un poids normatif, et indiquer qu'il serait préférable qu'un individu reçoive quelque chose de positif ou de négatif en échange de son mérite.

Jusque-là, il ne s'agit que de caractéristiques que nous avons déjà vues. Une telle perspective exclut l'idée que le mérite soit assimilable à l'ensemble de la moralité ou de la justice, mais n'exclut pas grand-chose d'autre. Olsaretti ajoute ensuite trois conditions dites « formelles », et deux conditions substantielles¹⁹. L'étude des conditions substantielles nous occupera plus tard ; examinons pour l'instant les trois conditions formelles proposées par Olsaretti. Celles-ci sont les suivantes : 1) le mérite ne doit pas être un mérite « moral », c'est-à-dire qu'il ne doit pas se fonder sur la vertu ou les qualités morales du caractère d'un individu. 2) Le mérite doit être pré-institutionnel, 3) le mérite doit posséder un fondement normatif indépendant, c'est-à-dire qu'il ne doit pas se réduire à un autre principe de justice.

Tous ces points ont fait l'objet d'un débat intense dans la littérature contemporaine sur le mérite. Comme tous sont importants, je vais les reprendre un à un dans les sections suivantes. Nous commencerons par le point qui est vraisemblablement le plus important : le mérite ne doit pas se réduire à d'autres principes de justice, et doit posséder un fondement normatif propre.

¹⁹ S. OLSARETTI, *Liberty, desert and the market*, op. cit., p. 24.

Cette remarque fondamentalement banale prend davantage de relief quand elle est opposée à une autre conception du mérite qui, bien que très populaire dans la littérature philosophique, fait perdre toute spécificité à l'idée de mérite : la conceptualisation du mérite comme compensation pour des souffrances ou des désavantages passés.

Le mérite comme compensation des souffrances passées

Pourquoi un concept utile de mérite doit-il posséder un fondement normatif propre ? Tout simplement parce que, si le mérite peut être réduit à un autre principe (comme la pure promotion de l'efficacité, ou de l'égalité), alors il n'est pas intéressant d'avoir un débat philosophique sur le mérite en tant que tel. Le véritable débat porterait alors sur le véritable fondement normatif en question. Si l'on imagine que le mérite peut être dérivé de principes comme l'efficacité ou l'égalité, alors, si l'on veut défendre le mérite, il faut avant tout défendre pourquoi l'efficacité ou l'égalité constituent d'authentiques valeurs. Si le mérite est purement dérivé, alors il n'y a plus vraiment de débats intéressants à avoir autour du mérite. Or, il existe de nombreuses théories du mérite qui ont, précisément, cherché à défendre le mérite, tout en le réduisant à un autre principe de justice. Parmi elles, la théorie du mérite comme compensation. Comme il s'agit d'une théorie assez importante dans la littérature, je vais commencer par faire une brève présentation du sujet, avant de montrer en quoi elle réduit entièrement la question du mérite à la question de la valeur de l'égalité.

Penser que le mérite est fondamentalement basé sur l'importance de compenser des souffrances, ou des dommages passés, est une conception extrêmement populaire dans la littérature philosophique. Elle a été par exemple défendue dans deux des rares ouvrages contemporains consacrés à la question du mérite, à savoir dans le *Desert* de George Sher, et dans *Giving Desert its Due*, de Sadurski²⁰. Sher défend cette conception dans le cadre de jugements de mérite liés à un travail passé ; Sadurski défend cette conception comme fondement général du mérite. De nombreux autres philosophes ont aussi, implicitement ou explicitement, vu dans la compensation le fondement philosophique du mérite. Par exemple, aussi bien Sher que Sadurski défendent l'idée qu'un travail mérite salaire parce que le travail représente une forme de désavantage, ou d'inconvénient, qu'il convient de compenser. Si un

²⁰ G. SHER, *Desert*, *op. cit.* ; Wojciech SADURSKI, *Giving Desert Its Due: Social Justice and Legal Theory*, Springer, 2010.

individu mérite une récompense, c'est avant tout pour compenser le fait qu'il aurait pu passer ces heures de travail à obtenir une plus grande satisfaction personnelle. Dans un cadre comparatif, si un individu a travaillé pendant qu'un autre se reposait, il est important de rétablir la balance et de récompenser l'individu travailleur pour qu'il regagne le niveau d'utilité globale dont bénéficie l'individu qui s'est reposé. Plus généralement, au-delà du cas du travail, un exemple courant est celui de la maladie ou de la malchance (comme dans l'exemple de Sher : « Gordon mérite d'avoir plus de chances à l'avenir ; il n'a pas eu de veine jusqu'à présent. »).

On peut distinguer deux problèmes vis-à-vis de la théorie du mérite comme compensation. Le premier problème est normatif, le second est conceptuel. D'un point de vue normatif, il n'est pas évident qu'il s'agisse d'une théorie particulièrement attirante en tant que modèle de société. Si le principal objectif des jugements de mérite est de compenser les souffrances, alors il n'y a pas de différence entre un travail foncièrement improductif et un travail extraordinairement productif, tant que les deux individus ayant fourni ces types de travail ont subi la même souffrance pendant leurs efforts. De plus, pour réaliser pleinement l'idéal du mérite comme compensation, il faudrait être en mesure d'évaluer à quel point les individus prennent plaisir à travailler, ou souffrent dans ce cadre. Par exemple, selon ce principe, un individu qui prendrait plaisir à travailler mériterait un salaire bien moindre qu'un individu qui détesterait son travail, et ce, même s'ils occupent exactement le même poste et travaillent le même nombre d'heures. Il est difficile de voir en quoi une théorie qui semble à première vue aussi contre-intuitive pourrait rendre compte des jugements ordinaires en faveur du mérite. Même si les défenseurs de l'idée de mérite comme compensation ont, de fait, cherché à reformuler leurs idées pour les rendre davantage compatibles avec les conceptions ordinaires, il n'en reste pas moins que cette théorie semble hautement discutable d'un point de vue normatif²¹.

Mais, si l'on veut prendre au sérieux l'idée que la compensation constituerait le fondement du mérite, le principal problème n'est probablement pas dans sa valeur normative. Le principal écueil est proprement conceptuel. L'idée du mérite comme compensation ne fournit pas une théorie du mérite, mais bien une théorie de l'égalité qui viserait à expliquer des jugements qui pourraient sembler être des jugements de mérite²². Ce n'est pas en soi une tare ; de fait, la

²¹ Pour un excellent résumé des défenses et des critiques de l'idée de mérite comme compensation, voir Owen MCLEOD, « Desert and Wages », *Utilitas*, 1996, vol. 8, n° 2, p. 205-221.

²² De fait, il existe au moins une théorie de la justice qui interprète l'idée de justice comme étant réductible à l'idée d'égalité, et qui considère que la justice est fondamentalement une forme de compensation. Il s'agit de la théorie proposée par Christopher Ake dans son article classique « Justice as Equality », *Philosophy & Public Affairs*, 1975, vol. 5, n° 1, p. 69-89. Ake y propose exactement le même type d'exemple que ceux offerts par des théoriciens du mérite comme compensation, tel Sher ou Sadurski. On voit bien que l'idée de mérite comme compensation se réduit fondamentalement à une théorie de l'égalité.

plupart des théories contemporaines de la justice visent à expliquer pourquoi les jugements apparents de mérite n'ont pas de valeur propre, mais sont en réalité dérivés d'un souci propre pour l'égalité ou bien le bien-être collectif. Le problème avec la théorie du mérite comme compensation est qu'elle prétend être une défense de l'idée de mérite, alors qu'il s'agit uniquement d'une réduction du mérite à l'idée d'égalité. Au lieu de fournir une clarification conceptuelle du mérite, il s'agit d'une accentuation de la confusion entre deux idées qui sont, et doivent rester, distinctes.

Tout d'abord, il s'agit de clarifier pourquoi le mérite comme compensation est avant tout une théorie de l'égalité, et non pas une théorie du mérite. Le principal problème est que cette théorie ne parvient pas à faire la distinction fondamentale entre compensation liée à une activité productive ou à un acte positif, et compensation due à une malchance pure. Cette théorie traite de la même façon un individu qui mérite compensation parce qu'il a subi un accident, et quelqu'un qui mérite une compensation pour une contribution à la société. Dans ce cadre, on voit bien pourquoi une telle conception échoue à proposer une analyse spécifique du mérite en tant que tel, c'est-à-dire en tant qu'il est compris dans les représentations quotidiennes. Comme je l'ai souligné plus haut, comprendre la spécificité du mérite implique de saisir qu'il s'agit d'un jugement lié à des actions ayant une valeur positive ou négative. Avoir subi un accident ne dit, en soi, rien de positif ou de négatif sur la personne en question. La théorie du mérite comme compensation renvoie uniquement à l'idée qu'il faut assurer l'égalité des individus au cours du temps, c'est-à-dire qu'il faut que les personnes obtiennent une quantité égale de biens ou de maux au cours de leur vie. Si un individu n'a pas obtenu les mêmes avantages que d'autres personnes à un moment donné, alors il doit obtenir une compensation à l'avenir pour obtenir une quantité de bien-être globale équivalente.

Il faut le répéter : il n'y a rien en soi de problématique à dire que le mérite n'a pas de consistance propre, et doit être réduit à d'autres valeurs comme à l'idée d'égalité. Cependant, une théorie qui se prétend être une théorie du mérite doit expliquer en quoi il existe une spécificité des jugements portant sur des actes positifs et négatifs. Défendre le mérite sous cette forme implique, de plus, d'admettre qu'il y existe une possibilité d'inégalités justes, à partir du moment où elles ont pour origine des différences de contribution. Étant donné que la théorie du mérite comme compensation écarte a priori la possibilité d'inégalités justifiées, elle ne peut être considérée comme une défense du mérite.

Pour mieux comprendre ce qui se joue ici, il faut faire appel à la différence entre une théorie philosophique conservatrice et une théorie révisionniste. Une théorie conservatrice cherche à rendre compte d'une idée morale présente dans la population, et de défendre le fait que cette

conception morale est globalement correcte. Une théorie révisionniste considère que la théorie morale en question est partiellement correcte, et partiellement erronée, et qu'il convient donc d'éliminer les parties erronées pour redéfinir cette conception. Dans le cas qui nous occupe, la théorie du mérite comme compensation se présente comme une théorie conservatrice du mérite, qui cherche à rendre compte des jugements de mérite tels qu'ils se trouvent dans la population. En réalité, il est évident qu'il s'agit d'une révision profonde des jugements de mérite : la théorie du mérite comme compensation développée par Sadurski, Sher, et d'autres, cherche avant tout à reformuler le mérite comme une théorie de l'égalité, en abandonnant les aspects les plus saillants du mérite (à savoir qu'il s'agit avant tout d'un jugement porté sur des actions positives, et non pas sur des accidents, ou des événements neutres d'un point de vue moral). Il est tout à fait possible de défendre une théorie révisionniste du mérite. Mais il serait prématuré de le faire avant d'évaluer dans quelle mesure les conceptions ordinaires du mérite sont défendables, ou même correctes. En tant qu'il cherche à défendre une conception déformée du mérite en la déguisant comme théorie des conceptions ordinaires du mérite, une théorie comme celle de Sadurski masque la nécessité de rendre compte des jugements quotidiens vis-à-vis du mérite.

L'idée qu'une conceptualisation intéressante du mérite doit reposer sur un fondement normatif indépendant peut sembler une idée tout à fait bénigne. Pourtant, comme nous venons de le voir, une telle conceptualisation permet de rejeter comme inadéquates un type important de théories du mérite, à savoir les théories du mérite comme compensation, qui considèrent que les jugements de mérite reposent fondamentalement sur une valorisation de l'égalité. Nous allons maintenant procéder à l'examen de la deuxième condition posée par Serena Olsaretti : l'idée que le mérite doit posséder une base pré-institutionnelle.

La conceptualisation rawlsienne, et les théories institutionnelles du mérite

Pour comprendre la deuxième condition posée par Olsaretti, à savoir qu'une conception viable du mérite doit être pré-institutionnelle, il faut revenir sur le long débat qui a opposé les théories pré-institutionnelles et institutionnelles du mérite²³. Une théorie institutionnelle du

²³ Un excellent résumé de ce débat se trouve dans Owen MCLEOD, « Desert and Institutions », in Louis P. POJMAN et Owen MCLEOD (dir.), *What Do We Deserve?: A Reader on Justice and Desert*, Oxford University Press, 1999, p. 186.

mérite est une théorie qui considère que les jugements de mérite ne prennent sens que dans le cadre d'un ensemble institutionnel qui possède une justification indépendante du principe de mérite lui-même.

Les théories institutionnalistes du mérite partagent toutes l'idée que le mérite n'est pas un principe de justice premier, et possède uniquement une valeur instrumentale. La valeur du mérite reposerait alors simplement dans le fait qu'il peut être un instrument dans la promotion de principes de justice qui, eux, posséderaient une valeur morale authentique. Dans ce cadre, le mérite ne prend un sens qu'au sein d'institutions qui ont été établies afin de mettre en œuvre ces principes de justice premiers. A l'inverse, les théories pré-institutionnelles du mérite sont celles qui défendent l'idée que le mérite est un principe de justice premier, qui ne saurait être dérivé d'autres principes supérieurs²⁴. Pour donner une idée claire de ce en quoi consiste une théorie institutionnelle du mérite, je vais présenter brièvement la théorie institutionnelle du mérite défendue par John Rawls dans sa *Théorie de la justice* et ses textes ultérieurs. Il s'agit en effet probablement de la théorie institutionnelle du mérite la plus développée, et qui a eu le plus d'influence par la suite. Par ailleurs, cette théorie est intéressante parce qu'elle accorde un contenu spécifique aux jugements de mérite, même si ce contenu est en fin de compte pleinement justifié à partir d'un appel à des jugements de justice extérieurs.

Même si Rawls rejette l'idée que le mérite puisse être un principe fondateur d'une théorie de la justice sociale, cela ne veut pourtant pas dire qu'il n'accorde aucune importance au mérite dans la théorie normative qu'il défend. Rawls cherche à montrer en quoi une reconceptualisation du mérite peut permettre de réconcilier les conceptions communes avec la théorie de la justice qu'il défend. Dans ce cadre, il propose donc une théorie révisionniste du mérite²⁵. Son interprétation repose sur l'idée que le mérite ne saurait être entendu que comme un principe subordonné, non fondateur, au sein d'une théorie de la justice. Cette idée a constitué l'acte fondateur des théories institutionnalistes du mérite.

Dans le cas de la théorie rawlsienne, le mérite est considéré comme dérivé des deux principes fondamentaux de justice. Rappelons brièvement le contenu de ceux-ci. D'après le premier

²⁴ Comme Samuel Scheffler l'a fait remarquer, il serait probablement plus approprié de décrire ces théories comme des théories pré-justicielles du mérite, puisqu'elles défendent non seulement l'idée que le mérite est antérieur d'un point de vue logique à l'existence d'institutions, mais qu'il est également indépendant de l'existence d'autres principes de justice. Cependant, par conformité avec la littérature existante, je continuerai de désigner les théories défendant une conception non-instrumentale du mérite comme étant des théories pré-institutionnelles. Samuel SCHEFFLER, « Distributive justice and economic desert », in Serena OLSARETTI (dir.), *Desert and Justice*, Oxford, Clarendon Press, 2007.

²⁵ La théorie de la justice reste souvent ambiguë vis-à-vis du fait de savoir dans quelle mesure Rawls propose une théorie conservatrice ou révisionniste du mérite. Si certains passages semblent montrer que Rawls cherche à réviser la conception commune, d'autres laissent entendre qu'il pense que sa reconceptualisation du mérite correspond effectivement aux conceptions ordinaires

principe de justice, dit principe d'égalité de liberté, « Chaque personne a une même prétention indéfectible à un système pleinement adéquat de libertés de base égales, qui soit compatible avec le même système de libertés pour tous²⁶. » Le second principe régit la justice de la distribution de positions sociales et d'avantages économiques. Il est constitué de deux parties, la première ayant priorité vis-à-vis de la seconde. La première partie du principe indique que « Les inégalités économiques et sociales [...] doivent d'abord être attachées à des fonctions et des positions ouvertes à tous dans des conditions d'égalité équitable des chances. » La seconde partie du principe indique que ces inégalités économiques et sociales doivent procurer le plus grand bénéfice aux membres les plus défavorisés de la société (le principe de différence)²⁷. »

Selon Rawls, ces deux principes de justice possèdent une valeur en soi. Ils sont l'expression de la liberté et de l'égalité des citoyens. Par exemple, le principe de différence est interprété comme la conséquence de l'idée de réciprocité appliquée à une société de citoyens coopérant mutuellement dans un système productif²⁸. A l'inverse, le mérite correspond à l'idée que les individus qui promeuvent les principes de justice doivent être récompensés, parce qu'une telle récompense assure le fait que les principes seront bien réalisés au maximum. Rawls considère donc que les individus méritants sont ceux qui promeuvent les deux principes fondamentaux de justice, et qui, notamment, améliorent ainsi la situation des plus défavorisés de la société. Il s'agit donc d'une conception purement instrumentale du mérite. Le mérite possède une force normative, mais une force normative dérivée, qui puise toute son importance des deux principes de justice fondamentaux.

Dans *La justice comme équité*, Rawls énonce trois interprétations possibles du mérite. La première interprétation est l'interprétation que Rawls considère comme l'interprétation intuitive du mérite, qu'il rejette comme principe politique de justice. Les deux autres interprétations sont des ré-interprétations révisionnistes de l'idée de mérite. Ces trois conceptions sont :

- D'abord, le mérite moral (*Moral Desert*), c'est-à-dire l'idée que la vertu et le caractère bon d'une personne appellent par eux-mêmes une récompense positive, et que la valeur de la personne dépend de ses qualités morales.

- Ensuite, les attentes légitimes (*Legitimate Expectations*). Les attentes légitimes traduisent l'idée que, au sein d'une société gouvernée par certains principes de justice, les membres de la société qui agissent conformément aux principes de justice acquièrent des titres (*Entitlements*)

²⁶ John RAWLS, *La justice comme équité*, La Découverte, 2008, sect. 13.

²⁷ *Ibid.*

²⁸ *Ibid.*, sect. 14.

vis-à-vis des biens qu'ils pouvaient s'attendre à recevoir s'ils agissaient en respectant les règles publiques.

- Enfin, le mérite institutionnel (*Deservingness*). Cette idée comble le manque que pourrait laisser l'idée d'attentes légitimes. En effet, même si certains individus ont agi conformément aux principes de justice et aux règles de la société, il y a des situations où le fait d'être récompensé pour ces actions n'est pas certain. Par exemple, dans un match de foot, même si les deux équipes jouent remarquablement bien et méritent toutes les deux en un sens de gagner, il ne peut y en avoir qu'une des deux qui remporte la partie ; l'équipe perdante est méritante, et aurait donc mérité de gagner, même si l'on ne peut rien faire pour changer cette situation²⁹.

Comme nous le verrons dans la suite de ce chapitre, pour Rawls, le mérite moral est une notion trop controversée pour faire partie d'une théorie de la justice. Seules les attentes légitimes et le mérite institutionnel sont repris au sein de sa théorie de la justice ; et, de ces deux conceptions, seules les attentes légitimes possèdent un poids normatif authentique : le mérite institutionnel traduit le fait qu'une situation est souhaitable, sans qu'il faille pour autant l'imposer. Les attentes légitimes constituent donc le cœur de la réinterprétation du mérite par Rawls.

La théorie du mérite défendue par Rawls est donc extrêmement sophistiquée, et propose une conception du mérite qui possède une certaine consistance. Contrairement à ce qu'il se passait dans le cadre des théories du mérite comme compensation, l'idée de mérite défendue par Rawls n'est pas complètement réductible à l'idée d'égalité ou de priorité. Le concept d'attentes légitimes conduit à une double justification de la récompense au mérite. La première justification est proximale, et tient à l'idée qu'il est bon de tenir les promesses faites aux citoyens. La deuxième justification est ultime, et se base sur l'idée qu'il est bon de créer un système d'incitations qui peut augmenter la richesse totale et par là améliorer la situation des plus démunis. La justification ultime réduit le mérite au rôle instrumental en faveur d'un autre principe de justice.

Les théories institutionnelles du mérite ne constituent donc pas entièrement une réduction des jugements de mérite à tout type de jugement servant une autre valeur. L'idée que l'État doit satisfaire les attentes légitimes des citoyens apporte un peu plus de consistance à la notion. Cependant, cela ne garantit qu'une spécificité et un intérêt limité aux jugements de mérite, qui sont fondamentalement dépendants d'autres principes de justice. En tant que telle, une

²⁹ Je résume ici la distinction faite par Rawls. Cf. *Ibid.*, p. 108.

conception institutionnelle du mérite n'accorde aucun rôle important au mérite en tant que principe de justice. Si une théorie veut défendre le mérite, et présenter une conception du mérite qui soit originale et qui ne se réduit pas ultimement à d'autres principes de justice, il faut adopter une conception non-institutionnelle du mérite. Le deuxième critère proposé par Serena Olsaretti s'en trouve donc justifié.

La troisième condition d'Olsaretti : la question du mérite comme récompense de la vertu

Nous pouvons passer maintenant à la troisième condition formelle exigée par Olsaretti, à savoir, le fait qu'une conception du mérite ne doit pas être une conception morale. Olsaretti ne veut pas dire par là que le mérite, pour être une idée intéressante, ne doit pas être une idée morale. Evidemment, dire que quelqu'un mérite quelque chose traduit une exigence morale. Olsaretti se sert du concept de mérite moral selon l'usage technique à l'œuvre dans la philosophie analytique : une conception du mérite est dite morale si elle considère que la vertu constitue le critère ultime du mérite.

Refuser une conception morale du mérite indique que je ne traiterai pas de ce qui a constitué l'objet d'étude d'une branche considérable de la littérature contemporaine sur le mérite : l'idée que les gens vertueux méritent d'être heureux, et davantage heureux que les personnes moins vertueuses, pour la simple raison qu'ils possèdent une disposition positive. Par exemple, à travers un certain nombre de livres ou d'articles au cours de ces vingt dernières années, des auteurs comme Ian Kershnar ou Shelly Kagan ont défendu l'idée que les individus vertueux méritaient d'être heureux, que les gens vicieux méritaient de souffrir, et qu'il devrait exister globalement une proportionnalité dans le bonheur des individus en fonction de leur valeur morale³⁰.

Même si une telle idée peut sembler quelque peu désuète, il s'agit d'une conception qui a eu une grande influence sur les débats portant sur le mérite au 20^e siècle. Des auteurs contemporains comme Kagan ou Kershnar poursuivent les intuitions défendues il y a plus de 80 ans par Ross dans *The Right and the Good*, quand celui-ci faisait appel à l'idée que la récompense de la vertu par le bonheur devrait faire partie d'une conception complète de la

³⁰ Shelly KAGAN, *The Geometry of Desert*, New York, Oxford University Press, 2015 ; Stephen KERSHNAR, *Desert and Virtue: A Theory of Intrinsic Value*, Lanham, Lexington Books, 2009.

justice³¹. Plus tard, dans les années 1960 et 1970, les critiques influentes du mérite portées par Friedrich Hayek et John Rawls ont été des critiques de cette idée d'une récompense du mérite en fonction de la vertu³². Par-delà leurs différences, les deux auteurs ont rejeté le mérite en grande partie du fait de leur perception de la profonde incompatibilité entre une théorie de la vertu et les valeurs portées par les sociétés libérales contemporaines.

Cette idée du mérite comme récompense de la vertu se distingue de trois façons par rapport au mérite tel qu'il est étudié dans cette thèse. Premièrement, il s'agit d'un mérite lié au caractère de la personne, et non à ses actions. Pour cette raison, j'appellerai ici un tel mérite mérite personnel, en opposition au mérite agentiel que j'étudierai dans ce livre. Le mérite personnel se distingue, de plus, par le fait qu'il présuppose un consensus sur la question de la moralité personnelle s'il veut se transformer en philosophie politique. Récompenser les individus en fonction de leur vertu est une activité qui peut sembler, à première vue, une tâche beaucoup plus complexe que la rémunération des actions.

Une telle conception de la justice repose sur deux aspects qui rendent toute traduction politique impossible, et qui expliquent pourquoi Rawls comme Hayek ont violemment critiqué la notion. Tout d'abord, l'idée de récompenser la vertu présuppose un consensus sur la valeur morale des individus qui n'existe plus du tout dans les sociétés libérales contemporaines. Pour déterminer dans quelle mesure un individu est vertueux, il semble nécessaire de faire appel à des conceptions de la vie bonne qui semblent tout à fait étrangères à l'idée d'une société libérale où les individus sont libres de mener la vie qu'ils souhaitent. Deuxièmement, il semble difficile d'imaginer comment il serait possible de mesurer la vertu des individus, et comment il serait possible de trouver un consensus politique à ce sujet. En dehors d'une connaissance divine, tout ordonnancement des individus à partir de leur caractère semble voué à l'échec. Dans ce cadre, si la théorie du mérite comme récompense de la vertu semble posséder une pertinence et une importance réelle d'un point théologique, il semble délicat de lui trouver des applications politiques possibles. Pour cette raison, l'étude ici sera essentiellement restreinte à la question du mérite pour des actions positives ou négatives, mais ne présupposant aucune conception spécifique de la vertu.

Cela ne veut évidemment pas dire que les ouvrages précédemment cités de Kershnar ou de Shelley Kagan ne comportent pas des analyses générales pertinentes sur la question du mérite, qui peuvent s'appliquer aussi bien à la question de la récompense des actions qu'à la récompense de la vertu proprement dite. Effectuer un parallèle entre la récompense des actions

³¹ D. ROSS, *The Right and the Good*, *op. cit.*

³² J. RAWLS, *Théorie de la justice*, *op. cit.* ; F. A. HAYEK, *Law, Legislation and Liberty*, *op. cit.*

et la récompense du caractère individuel peut conduire à une meilleure compréhension de l'idée de mérite en général. Cependant, afin de conserver la question d'une possible application politique des analyses présentées ici, tant que cela n'est pas précisé, les jugements de mérite seront ici entendus comme jugements portant sur la question de la récompense possible ou souhaitable d'actions déterminées.

Pauvre méritant ou non-méritant : brève analyse d'une conception courante

Avant d'abandonner définitivement l'étude de cette question, il est important de considérer un dernier aspect lié à la question du mérite personnel et à la vertu d'un individu. Dans les débats politiques apparaît souvent l'idée de pauvres méritants, d'étudiants méritants, ou de personnes méritantes en général. Or, cette idée d'être méritant semble bien renvoyer à un certain type de vertu, ou à un certain type d'excellence de caractère en général. On peut faire alors la distinction entre deux usages de l'idée de mérite dans le langage courant. Si l'on parle d'une personne « méritante », alors l'idée de mérite semble renvoyer essentiellement à la vertu ou au caractère d'un individu. Lorsqu'on parle du fait que quelqu'un « mérite » quelque chose, alors l'idée de mérite renvoie en grande partie au fait que cette personne a fait quelque chose de positif ou négatif, et qu'il est bon qu'elle reçoive une récompense ou une punition en retour. Dans un cas il y a un jugement d'excellence, dans l'autre il y a l'idée d'une action possédant une valence positive ou négative. Les deux types de jugement sont extrêmement courants dans le langage quotidien. Ils renvoient cependant à des idées morales distinctes.

Un exemple amusant de la distinction entre l'idée d'être méritant, et l'idée de mériter quelque chose, se trouve dans un souvenir raconté par Jean-Paul Sartre dans *Les mots*. Sartre y écrit à propos de sa gouvernante : « M^{lle} Marie Louise me démoralisait. Je croyais les salaires proportionnés au mérite et on me disait qu'elle était méritante : pourquoi donc la payait-on si mal³³ ? » Dans cette citation, l'idée que Marie Louise était méritante renvoie à l'idée qu'elle représente une certaine excellence morale. Cependant, pour la plupart des adultes, cela ne veut pas dire qu'elle mérite d'être très bien payée, étant donné que la contribution socio-économique effectuée par quelqu'un ne correspond pas forcément à l'excellence de son caractère moral.

³³ Jean-Paul SARTRE, *Les mots*, Paris, Gallimard, 2009, p. 66.

Sartre joue ici de la ressemblance entre deux termes apparentés pour mettre en avant l'indignation de son double enfant.

Comme le lecteur l'aura compris, je me concentrerai dans l'ensemble de cette thèse sur la question de la justification de l'idée que des individus méritent quelque chose pour leurs actions bonnes ou mauvaises. En d'autres termes, je ne me concentrerai pas sur l'idée de savoir ce qu'est une personne de mérite (au sens d'une personne vertueuse), ni sur l'idée de ce qu'est une personne méritante. Il est temps de lever une ambiguïté dans l'usage du vocabulaire tournant autour de l'idée de mérite, qui avait conservé un certain flou jusqu'à présent. On peut parler du mérite d'un individu pour parler du caractère excellent de cette personne. Cependant, quand je parle de jugement de mérite, je veux parler du type de jugement qui indique que quelqu'un mérite quelque chose parce qu'il a fait une bonne ou une mauvaise action. L'idée de jugement de mérite ne renvoie donc pas au fait de jauger de l'excellence d'un caractère. De la même manière, l'idée de personne méritante peut renvoyer à deux choses. Elle peut renvoyer, dans l'acception courante, à l'idée qu'une personne possède une forte vertu, ou un caractère excellent. En dehors de cette section, et plus généralement au cours de cette thèse, l'idée de personne méritante renverra à un raccourci pour exprimer l'idée que cette personne mérite une récompense parce qu'elle a effectué une contribution positive. Même s'il y a un risque de confusion avec l'idée de personne méritante comme personne vertueuse, parler de personne méritante pour parler d'une personne méritant une récompense nous fera gagner beaucoup de temps.

Les réflexions qui nous occuperont pendant cette thèse seront donc tout autres que celles liées au caractère méritant des individus, entendu au sens de caractère vertueux. Cependant, du fait du caractère extrêmement courant de l'idée d'une personne méritante, je vais procéder à une brève analyse de la manière dont le concept est utilisé dans le langage ordinaire.

Dans le cas du pauvre méritant, le terme semble renvoyer à un individu qui ferait preuve d'un véritable désir de contribuer au bien-être de la communauté, en opposition au pauvre non-méritant qui voudrait simplement profiter d'un soutien public sans apporter sa part à la coopération sociale³⁴. La vertu dont fait preuve le pauvre méritant est celle d'être un bon coopérateur, prêt à participer à un système généralisé de production commune. Dans le cas de la personne méritante en général, la vertu en question semble être la capacité à persévérer

³⁴ Il va sans dire que je cherche ici simplement à analyser le contenu de conceptions sociales courantes, sans que cela n'engage en rien en quoi ces conceptions renvoient à une description pertinente de la réalité sociale. L'idée courante de pauvres non-méritants est en soi intéressante à étudier pour ce qu'elle révèle des conceptions morales et sociologiques ordinaires, sans qu'il soit nécessaire de juger si le concept de pauvre non-méritant correspond à une partie négligeable ou importante des membres de la société.

dans des efforts possédant un but positif, malgré des obstacles importants³⁵. Un étudiant méritant est un étudiant qui a fait face à un certain nombre de difficultés, notamment parce qu'il ne possédait pas les avantages sociaux liés au fait d'être né dans une famille socialement favorisée, et qui néanmoins a persévéré et a tenté de réussir quand même ses études.

Ces deux questions, celle du pauvre méritant et de l'étudiant méritant, possèdent une valeur politique évidente. Il semble néanmoins que ces deux questions renvoient à une analyse philosophique différente de celle liée à la justification des inégalités méritées. Dans le cas du pauvre méritant, il est très facile de savoir ce qu'il mérite : il mérite d'être aidé, jusqu'à ce qu'il sorte de sa situation de pauvreté. Ceci n'est pas lié au fait qu'il posséderait une qualité qui le mettrait à part des autres citoyens ; au contraire, le fait qu'il est prêt à participer à la coopération sociale le rend simplement l'égal de tous les autres citoyens, un membre à part entière de la communauté productive. Dans les conceptions ordinaires, le fait qu'il est prêt à travailler ou à contribuer sert à compenser le fait qu'il ne travaille pas ou ne contribue pas effectivement ; il traduit le fait que, dans un monde contrefactuel où le pauvre méritant aurait la possibilité de travailler, il le ferait et serait un contributeur à part entière. La conception de pauvre méritant renvoie donc fondamentalement à la question de la fondation de la communauté de justice : celui qui est prêt à travailler doit pouvoir bénéficier du système de réciprocité généré par la société. La question du pauvre méritant renvoie donc à un cas particulier des obligations de réciprocité générées par la coopération sociale. Justifier une analyse en termes de pauvres méritants ou non-méritants ne nécessite donc pas nécessairement une justification sophistiquée en termes de mérite ; toute théorie fondée sur un système de coopération et de réciprocité est à même de faire la même distinction. Même si l'on considère que tous les travailleurs sont fondamentalement également méritants, ou que le mérite n'est pas un principe de justice fondamental, on peut effectuer cette distinction. Des théories de la justice fondées sur le principe de réciprocité développées par John Rawls ou Elizabeth Anderson sont à même de faire cette distinction entre pauvre méritant et non-méritant, même si ces théories ne s'appesantissent pas en général sur cette question.

Dans le cas de l'étudiant méritant, il est également facile de dire ce qu'il mérite : il mérite de réussir. Dans ce cadre, une réponse politique adaptée reviendrait à mettre en place un système social qui limite les obstacles rencontrés par les individus, afin que seuls des facteurs pertinents, comme les efforts ou (selon certaines conceptions) les talents, aient une influence décisive sur le destin des individus. Dans ce cadre, la question d'un étudiant méritant renvoie bien sûr à la

³⁵ C'est par exemple l'analyse que faisait déjà Renouvier en 1869, dans son classique *Science de la morale*, Paris, Fayard, 2002, chap. XXXVIII.

question de l'égalité des chances, et à la possibilité de créer des structures qui minimisent l'existence de facteurs non pertinents pour la réussite des individus. Un examen plus poussé de la question de l'égalité des chances attendra la dernière partie de cette thèse.

Nous avons donc fait l'analyse de deux cas très particuliers, où un jugement de mérite semble effectivement se fonder sur un certain trait de vertu, ou tout du moins sur le fait qu'un individu possède une certaine qualité. Il semble cependant assez clair que ces cas n'impliquent pas qu'une théorie correcte du mérite soit une théorie fondée sur la nécessité de récompenser l'excellence morale en général. Il n'y a pas de lien évident entre l'idée qu'un pauvre méritant doit être secouru, ou qu'un étudiant méritant doit réussir, et l'idée que les saints devraient être plus heureux que le commun des mortels. Parce que ce dernier type de jugement semble renvoyer à un ensemble de conceptions potentiellement assez éloignées du monde politique, et pour limiter l'ensemble du débat, je me limiterai donc dans le reste de cette thèse à une simple étude des jugements de type : X mérite une récompense (ou une punition) parce qu'il a effectué une action de type positif (ou négatif).

Les deux critères substantiels d'Olsaretti : égalité des chances et justice comparative

Nous avons donc vu les différents critères formels proposés par Olsaretti : une conceptualisation intéressante du mérite doit reposer sur un fondement normatif indépendant, doit être pré-institutionnelle, et ne doit pas reposer sur la récompense de la vertu d'un individu. Avant cela, nous avons également vu que le mérite doit correspondre à la récompense d'une action valorisée.

Passons maintenant aux deux conditions substantielles proposées par Olsaretti. Tout d'abord, pour Olsaretti, si l'on veut qu'une conception du mérite soit plausible, il faut qu'elle présuppose l'existence d'une juste égalité des chances d'acquérir du mérite. Ensuite, selon Olsaretti, le mérite doit être compatible avec les règles de la justice comparative. Je vais d'abord expliquer brièvement la première condition, et dire pourquoi elle ne me semble en rien évidente. J'expliquerai et défendrai ensuite la deuxième condition.

Olsaretti considère donc qu'il doit exister une juste égalité des chances d'acquérir le mérite. La philosophe veut dire par là qu'une conception du mérite qui accepterait qu'on puisse être méritant sans que tous aient eu les mêmes opportunités d'être méritants ne fait pas de sens d'un

point de vue moral. Prenons un exemple concret. Si un ingénieur construit un pont, ou fait un travail de grande qualité, Olsaretti considère qu'on ne peut pas dire qu'il « mérite » un meilleur salaire qu'un ouvrier, sauf si cet ouvrier disposait à l'origine d'exactly les mêmes chances de devenir ingénieur. Sans une juste égalité des chances de contribuer à la société, récompenser les contributions sociales est injuste. Olsaretti parle de « juste » égalité des chances pour reprendre le concept rawlsien et par là indiquer qu'il ne s'agit pas d'une simple égalité des chances formelles, ou de carrière ouverte aux talents. Elle a donc en tête l'idée que tous les individus doivent avoir les mêmes opportunités dans tous les milieux sociaux.

Une telle condition peut sembler plausible au premier abord. Cependant, elle n'a fondamentalement rien d'évident. Des contre-exemples à cette exigence sont connus de longue date. Le cas le plus cité prend l'exemple d'une personne qui vient de sauver une deuxième personne qui se noyait dans un lac. Le sauveteur mérite-il les remerciements de la personne sauvée ? Si l'on imagine que le sauveteur en question passait par hasard au bord du lac, et qu'il n'y avait pas de raison de penser qu'une personne se noierait dans le lac à ce moment-là, alors il n'existait pas de juste égalité des chances de sauver la personne se noyant dans le lac. Le sauveteur disposait d'une position privilégiée, dont il n'était nullement responsable, pour se rendre méritant. La condition posée par Olsaretti exigerait donc que le sauveteur ne mériterait pas plus de remerciements que n'importe quelle autre personne qui ne se trouvait pas au bord du lac à ce moment-là. Mais une telle condition peut sembler absurde.

Bien sûr, la question du lien entre mérite et responsabilité exige une plus longue explication, et le critère proposé par Olsaretti n'est pas absurde au point qu'il soit possible de le rejeter sans plus de procès. Cependant, cette discussion justifie le fait que je n'accepte pas la première condition substantielle d'Olsaretti, et que je repousse la discussion du lien éventuel entre mérite et responsabilité à la quatrième partie de cette thèse.

La deuxième condition substantielle posée par Olsaretti est moins controversée. Cette condition exige que les jugements de mérite respectent les principes de la justice comparative. Cela signifie que, si un individu a effectué une action qui est reconnue comme étant une plus grande contribution que l'action effectuée par un autre individu, alors le premier individu mérite une plus grande récompense que le second ou, du moins, une récompense qui n'est pas inférieure. Il s'agit, en d'autres termes, d'une condition d'ordre : il doit y avoir un ordre similaire entre les jugements de mérite, d'un côté, et les jugements de récompense, de l'autre. Une théorie du mérite ne serait pas plausible si elle autorisait le fait qu'une action de peu de mérite corresponde à une plus grande récompense qu'une action de grand mérite. Cette deuxième conception correspond à l'idée populaire selon laquelle les récompenses doivent être

proportionnées aux contributions. Si l'on abandonne l'idée d'une proportionnalité mathématique exacte en faveur d'un certain ordre, plus flou et plus compatible avec le caractère vague de nombreux jugements de mérite, alors on obtient une conception plausible.

Mériter une récompense, mériter une peine

Avant de clore ce chapitre, je tiens à préciser que, dans le cadre de cette thèse, je traiterai de la même manière le fait de mériter un salaire et de mériter une punition. Ce traitement parallèle des deux idées se fera presque toujours de façon implicite. Ainsi, la plupart du temps, pour simplifier l'exposition, je parlerai simplement de « mériter une récompense » sans parler de manière similaire du fait de mériter une punition, même si mon analyse s'appliquerait de la même manière aux deux questions.

D'une certaine manière, ce traitement parallèle du fait de mériter une peine et de mériter une récompense ne demande pas de justification particulière. De nombreuses théories classiques du mérite, d'Adam Smith à Joel Feinberg, ont traité les deux notions de manière similaire sans pour autant considérer que ce parallèle demandait justification, tant cela leur semblait évident³⁶. Je pense effectivement que les deux idées de mérite positif et négatif sont si proches qu'il est parfaitement justifié de les traiter de façon identique.

Repassons en revue les différentes caractéristiques des jugements de mérite que nous avons étudiées dans ce chapitre. Dans le cas d'un mérite positif comme négatif, un jugement de mérite correspond à l'idée qu'une personne doit recevoir quelque chose possédant une certaine valence (positive ou négative) ; cela est dû au fait que cette personne a effectué une action possédant une certaine valence ; et il doit y avoir une correspondance entre la valence de l'action effectuée et la valence de ce que reçoit la personne. Autrement dit, plus une personne a fait quelque chose de mal, et plus cette personne doit être punie, et plus elle a fait quelque chose de bien, et plus elle doit être récompensée.

Dans les deux cas, le mérite peut satisfaire les conditions formelles posées par Olsaretti. Dans le cas du mérite positif comme négatif, le mérite peut être conçu ou bien de façon dérivée, ou bien de façon première. Il est tout à fait possible d'avoir une théorie du mérite négatif qui réduise la punition à un autre principe de justice, comme l'égalité, ou l'efficacité. Les théories

³⁶ Adam SMITH, *Théorie des sentiments moraux*, PUF, 2014 ; J. FEINBERG, *Doing and Deserving*, *op. cit.*

utilitaristes classiques réduisent généralement l'idée qu'un criminel mérite punition au fait que c'est un pur moyen pour prévenir de futurs crimes. Dans un autre contexte philosophique, la théorie de la peine comme « compensation » inversée est une théorie pénale qui fonde le mérite négatif sur le fait que punir les criminels sert l'égalité entre les individus. Cette théorie correspond à l'idée qu'en rompant la loi, le criminel a profité d'un bien indu, et que ce bien indu doit être compensé par une punition³⁷. Dans ce cas, on a une réduction du mérite pénal à une théorie de l'égalité au cours du temps, exactement de la même manière que la théorie de la récompense comme compensation réduisait le fait de mériter un salaire à une théorie de l'égalité.

De plus, il existe des théories institutionnelles du mérite négatif. La théorie pénale de Hart en est un exemple frappant³⁸. Hart considère que la peine est justifiée par un cadre institutionnel général qui sert à limiter le nombre de crimes au sein de la société. La justification est donc foncièrement dissuasive. Cependant, la justification proximale de la peine (ce que Hart nomme les règles de distribution de la peine) fait appel à l'idée qu'un citoyen est responsable de ce qu'il a fait, et que, ayant été prévenu par l'État qu'il ne devait pas enfreindre les lois, il n'a pas de raison de se plaindre d'avoir été puni. On retrouve la même combinaison de critères que dans la théorie de la justice de Rawls : une justification ultime en termes d'une autre valeur (le principe de différence chez Rawls, la dissuasion chez Hart) et une justification proximale en termes d'un autre critère, apparenté au mérite, mais légèrement différent (la satisfaction des promesses chez Rawls, la responsabilité chez Hart).

Enfin, le mérite positif comme négatif doit satisfaire aux règles d'une justice comparative, c'est-à-dire que, si un individu a effectué un crime plus grave qu'une autre personne, alors la personne ayant effectué le crime le plus grave doit être davantage punie que l'autre personne.

Dans tous ces cas, il existe un parallèle parfait entre le mérite négatif et positif. Je vois mal pourquoi les deux notions exigeraient un traitement séparé. Enfin, comme nous allons le voir dans les chapitres 6 à 9, les deux idées peuvent être justifiées en faisant appel au même type de raisons. Je défendrai donc une théorie de la peine comme de la récompense qui satisfait les conditions posées par Olsaretti, c'est-à-dire une théorie qui n'est pas fondée sur un autre principe de justice, qui n'est pas institutionnelle, qui n'est pas fondée sur la vertu, et qui satisfait les principes de justice comparative.

³⁷ C'est la théorie de la peine que propose Georges Sher. G. SHER, *Desert, op. cit.*

³⁸ H. L. A. HART, *Punishment and Responsibility: Essays in the Philosophy of Law*, Oxford, Oxford University Press, 2008 ; Bertrand GUILLARME, *Penser la peine*, Paris, Presses Universitaires France, 2003.

Bien sûr, il existe un obstacle linguistique au fait de traiter les deux notions à l'identique. Le terme de « mérite », en Français comme dans d'autres langues, possède de fortes connotations positives. On ne parle jamais du « mérite » d'un criminel, sauf si l'on souhaite indiquer que cette personne a, malgré son crime, effectué de bonnes actions. Cependant, ces habitudes linguistiques ne valent pas lieu d'analyse conceptuelle. Comme nous venons de le voir, l'idée de mérite à l'œuvre dans l'idée de récompense et de punition trahit une très forte similarité conceptuelle. De plus, il existe un usage parallèle du verbe « mériter » dans « mériter une récompense » et « mériter une peine », et cet usage me semble refléter une parenté profonde entre les deux idées.

Pour cette raison, l'ensemble des analyses de mérite à l'œuvre dans cette thèse s'appliquera de la même manière à l'idée de mérite positif et négatif. Si l'idée de mérite négatif n'est pas mentionnée, c'est uniquement pour simplifier l'exposition. A l'occasion, je mentionnerai bien séparément les deux idées, par exemple dans le chapitre 7, consacré à la théorie communicationnelle de la peine. Mais même dans ce cas-là, comme je le montrerai, les mêmes analyses et les mêmes justifications peuvent s'appliquer aux mérites positif et négatif.

Valorisation du mérite et récompense du mérite dans la vie réelle

Avant de clore ce chapitre, il nous reste à étudier une confusion particulièrement problématique dans les discours courants sur le mérite. Je vais ainsi examiner une critique de l'idée de mérite qui est relativement rare chez les philosophes, mais qui me semble extrêmement courante chez dans la population en général.

En un mot, en général, la plupart des critiques du mérite sont en réalité des critiques d'une société où le mérite n'est pas récompensé. Il s'agit donc de critiques sociales qui dénoncent une société injuste au nom des principes mêmes de mérite. Par exemple, le fait de mettre en avant que l'égalité des chances n'est pas réalisée au sein de notre société est souvent présenté comme étant une critique du mérite. En réalité, il est évident que si l'on considère que l'absence d'égalité des chances est une tare sociale, alors on adopte en réalité des principes méritocratiques pour lesquels une société idéale serait une société où régnerait l'égalité des chances.

Cette thèse constitue une défense de l'idée de mérite, mais il est évident que je ne ferai à aucun moment l'hypothèse que notre société est une société parfaitement juste et

méritocratique. De la même manière, il est tout à fait possible de défendre l'idée d'égalité sans considérer que la société actuelle est parfaitement égalitaire. Le factuel et le normatif représentent deux ordres séparés. C'est une évidence, mais une évidence qui est malheureusement mal comprise par de nombreux critiques du mérite. Les critiques philosophiques maintiennent généralement la distinction entre le normatif et le factuel, et leur critique du mérite est donc la plupart du temps une critique authentique du mérite, et non pas une critique du manque de réalisation du mérite au sein de la société. Les discours courants ne sont malheureusement pas toujours aussi prudents ; et ce manque de prudence se retrouve parfois également chez quelques auteurs en sciences sociales.

Plus particulièrement, des auteurs comme François Dubet ou Marie Duru-Bellat ont présenté ces dernières années une critique du mérite qui prend trois aspects³⁹. Tout d'abord, ils montrent, de façon convaincante, que l'origine sociale a encore un impact déterminant dans les sociétés actuelles, et que les idéaux méritocratiques ne sont pas appliqués. Comme nous venons de le voir, un argument de ce genre n'est pas vraiment une critique du mérite, mais constitue plutôt une défense implicite de l'idéal de mérite comme moteur d'une critique des injustices de nos sociétés.

Ensuite, Dubet ou Duru-Bellat défendent l'idée que les discours sur le mérite peuvent être exploités pour justifier des inégalités sociales qui sont en réalité injustifiables. Bien sûr, il s'agit d'un risque. Mais le fait qu'un idéal moral peut être détourné n'est en aucun cas un argument contre l'usage correct de cet idéal. Le fait que l'on ait pu mobiliser l'idée d'égalité pour critiquer les libertés formelles jugées « bourgeoises » n'est en aucun cas un argument contre un usage correct de l'idée d'égalité. Le fait que les libertariens peuvent rejeter l'égalité au nom de la liberté ne montre en aucun cas que l'idée de liberté est fondamentalement viciée. Le fait que des mouvements politiques dont on peut réprocher les orientations se sont servis de l'idée de mérite ne montre pas qu'il s'agit d'un idéal néfaste. Toute la question est de savoir comment trouver un équilibre entre les différents idéaux moraux comme l'égalité, la liberté ou le mérite, tout en dénonçant les discours mobilisant de façon idéologique ces mêmes idées.

Enfin, troisièmement, François Dubet et Marie Duru-Bellat critiquent souvent une focalisation sur le mérite comme seul principe de justice sociale. Ils soulignent notamment l'importance de considérer une pluralité de principes de justice. Même s'il s'agit d'une idée avec laquelle je suis tout à fait d'accord, il ne s'agit pas vraiment d'une critique du mérite en tant que tel, mais une critique de visions de la justice exclusivement centrées sur le mérite.

³⁹ François DUBET, *Les places et les chances : Repenser la justice sociale*, Paris, Le Seuil, 2010 ; Marie DURU-BELLAT, *Le mérite contre la justice*, Les Presses de Sciences Po, 2009.

Je ne suis pas certain qu'il soit utile de s'aventurer dans une discussion étendue de la mise en œuvre du mérite au sein de nos sociétés dès le début de cette thèse, mais je soulignerai dès à présent que je considère qu'une application correcte du mérite ne constituerait en aucun cas une justification des inégalités dans nos sociétés. Comme l'a bien mis en avant Thomas Piketty et de nombreux autres économistes, une grande partie des inégalités de richesse dans la société actuelle est le fruit de l'héritage, ou de la reproduction du capital⁴⁰. Ces inégalités ne peuvent en aucun cas se justifier par le mérite, ou par les talents ou les efforts supérieurs des individus concernés. De façon générale, les deux seules manières de « justifier » les inégalités actuelles sont, il me semble, ou bien de les justifier par des théories économiques qui soulignent l'impact positif en termes d'efficacité, ou bien de les justifier par une application absolue du droit de propriété, à la façon libertarienne. A l'inverse, une application stricte du mérite impliquerait une remise en cause radicale du fonctionnement de nos sociétés, au moins par le biais d'une importance bien moindre accordée à l'héritage et une taxation bien plus forte des revenus du capital. Les auteurs libertariens ou néo-libéraux ne s'y sont pas trompés, en proposant une critique radicale du mérite, à la façon de Hayek, ou en n'accordant aucune importance au mérite dans leur théorie, à la façon de Nozick⁴¹.

Le lecteur pourra donc garder en tête que, quand j'indique qu'une théorie du mérite peut justifier des inégalités, je ne fais pas référence à la justification des inégalités telles qu'elles se trouvent dans la société actuelle. La différence de revenus justifiée par des différences de productivité est une question épineuse. Cependant, il me semble légèrement aberrant de supposer que des différences de salaire d'un à mille, et des différences de revenu d'un à cent mille, telles qu'on peut les trouver dans nos sociétés, puissent être justifiées par des différences de mérite. Dans la société actuelle, quelle que soit l'interprétation qu'on en fasse, le principe de mérite est une idée critique.

Conclusion

Ce chapitre a essentiellement cherché à limiter la conception du mérite, pour éviter de l'assimiler à des conceptions étrangères comme l'égalité ou la compensation. Nous avons défini

⁴⁰ Thomas PIKETTY, *Le Capital au XXIe siècle*, Paris, Le Seuil, 2013.

⁴¹ F. A. HAYEK, *Law, Legislation and Liberty*, op. cit. ; Robert NOZICK, *Anarchy, State and Utopia*, New York, Basic Books Inc., 1974.

un jugement de mérite comme tout type de jugement impliquant qu'il serait juste qu'une personne reçoive une récompense ou une punition du fait d'un trait positif ou négatif propre à cette personne. Ce trait positif peut, essentiellement, correspondre ou à une action, ou à un trait de caractère. Pour des raisons liées à la nécessité de concevoir une philosophie politique publique, nous avons décidé de nous concentrer ici sur les jugements de mérite liés à des actions, et non pas au caractère vertueux ou vicieux des personnes.

Nous avons également, à la suite des réflexions de Serena Olsaretti, proposé une série de conditions permettant de caractériser une conception intéressante du mérite. Parmi ces conditions, la plus importante est sans doute l'idée que le mérite doit posséder un fondement normatif propre, et ne doit pas être dérivé d'autres principes fondamentaux. Une deuxième condition que nous avons reprise correspond à l'idée que le mérite doit respecter les principes d'une justice comparative, c'est-à-dire qu'il doit être possible d'ordonner le mérite des individus.

Une telle série de conditions peut sembler extrêmement anodine. En réalité, elle permet d'évacuer une grande partie des conceptions contemporaines du mérite, comme ne présentant pas un concept suffisamment pertinent du mérite. Les conceptions du mérite proposées par Fred Feldman, Owen McLeod ou Geoff Cuprit sont rejetées parce qu'assimilant le mérite à la justice toute entière, et fournissant ainsi une conception trop large du mérite⁴². L'idée de Georges Sher ou de Sadurski selon laquelle le mérite consiste dans la compensation des désavantages subis est rejetée comme défense du mérite, car il s'agit fondamentalement d'une réduction du mérite à l'égalité. Les théories institutionnelles du mérite, comme celle défendue par John Rawls, sont rejetées comme réduisant le mérite à un instrument au service d'autres valeurs ou d'autres principes de justice⁴³. Les théories du mérite comme vertu, comme celles défendues par Kershnar ou Kagan, sont rejetées car ne proposant aucune perspective politique claire.

Il sera bien sûr possible de revenir par la suite à une de ces conceptions. Il est possible que nous arrivions à la conclusion que les jugements de mérite « spécifiques », ceux qui peuvent en apparence justifier les inégalités du fait de l'inégalité de contribution, sont en réalité invalides, et que seuls des jugements de compensation, fondés sur l'égalité, sont valides. Il est également possible que toute conception plausible du mérite soit en fin de compte institutionnelle. Si cela

⁴² F. FELDMAN, *Distributive Justice*, *op. cit.* ; O. MCLEOD, « Desert and Wages », *op. cit.* ; Geoffrey CUPIT, *Justice As Fittingness*, Oxford, Oxford University Press, 1997.

⁴³ Il faut bien noter qu'il ne s'agit pas là d'une critique de Rawls, ou des théories institutionnelles du mérite ; ces auteurs savent très bien que leurs théories n'offrent qu'une place subordonnée et limitée à l'idée de mérite. Il s'agit bien, cependant, d'une critique des théories de Feldman, McLeod ou Sadurski : ces auteurs prétendent analyser un concept original, celui de mérite. Mais le fait que ce concept « original » se réduit aux concepts de justice ou d'égalité fait perdre tout intérêt à une analyse du « mérite » en tant que concept possédant une valeur propre.

est vrai, alors notre enquête aura fini par rejeter toute idée de mérite authentique. L'examen des justifications possibles du mérite aura lieu dans la troisième partie de cette thèse. Dans l'immédiat, il s'agira de rendre un peu plus concrète l'idée de « contribution » que nous avons manipulée jusqu'ici de façon intuitive, pour mettre en avant une théorie précise des critères qui peuvent rendre un individu méritant. Cet examen des critères régissant la distribution des récompenses au mérite nous occupera dans notre chapitre 2.

Chapitre 2

Les critères du mérite

Dans ce chapitre, j'examinerai la question des critères nécessaires pour attribuer du mérite à quelqu'un. Il s'agit d'une question classique en philosophie politique et dans le débat public, l'importance relative de l'intention, des efforts, et de la contribution ayant fait l'objet de débats sophistiqués déjà chez Thomas d'Aquin, Adam Smith, ou Kant⁴⁴. Parmi les principaux critères, le débat autour de l'importance respective des efforts et de la contribution a été particulièrement virulent. Il s'agit de l'un des points des plus controversés, au point que John Stuart Mill, dans *l'Utilitarisme*, a défendu l'idée que l'impossibilité de choisir entre la contribution et les efforts pour déterminer du mérite d'un individu impliquait qu'il fallait choisir une autre théorie morale, utilitariste, et renoncer aux théories morales du sens commun⁴⁵.

La question du critère du mérite est distincte de la justification du mérite. Pour prendre un exemple simple, imaginons qu'un individu ait contribué positivement à une entreprise, et que l'on dise qu'il mérite de recevoir une récompense pour cela. Dans ce cas précis, le critère du mérite est le fait d'avoir fait une contribution positive. Mais cela n'indique en rien quelle est la justification de ce jugement, c'est-à-dire que cela ne donne aucune information sur les raisons pour lesquelles récompenser cet individu serait justifié. Même si différentes théories s'accordent sur l'idée que les individus doivent être récompensés pour leur contribution, elles peuvent le faire pour des raisons totalement différentes : une théorie pourrait justifier ce critère de mérite par l'idée qu'il faut stimuler la production, tandis qu'une autre théorie pourrait justifier ce critère avec l'idée qu'il est important de récompenser les individus vertueux, indépendamment de l'impact productif d'une telle récompense. La question de la justification est complètement indépendante de la question du critère du mérite.

⁴⁴ THOMAS, *Textes sur la morale, op. cit.* ; Adam SMITH, *The Theory of Moral Sentiments*, New York, N.Y., Penguin Classics, 2010 ; I. KANT, *Fondements de la métaphysique des mœurs, op. cit.*

⁴⁵ John Stuart MILL, *L'utilitarisme*, trad. fr. Georges TANESSE, Paris, Flammarion, 2010.

Cette distinction recoupe la distinction faite par Hart entre la définition, le but justificatif général, et la distribution d'un bien ou d'un mal. Ainsi qu'indiqué par Hart dans ses « Prolégomènes à une théorie de la peine⁴⁶ », « Dans le cas de la propriété nous devons distinguer la question de la définition de la propriété, la question de savoir pourquoi et dans quelles circonstances il est bon de maintenir cette institution, et les questions de savoir dans quelle mesure des individus peuvent avoir le droit d'acquérir de la propriété et quelle quantité ils devraient avoir le droit d'acquérir. Ces questions peuvent être nommés les questions de Définition, de But Justificatif Général, et de Distribution, cette dernière question étant divisée en question de Titre et de Quantité. » Ces différentes questions essentielles se posent également dans le cas du mérite. En ce qui concerne la définition de la distribution conforme au mérite, on peut indiquer qu'il s'agit d'un type de distribution où l'on attribue à un individu des biens ou des maux qui doivent correspondre à une action positive ou négative causée par l'individu (comme nous l'avons vu au cours du chapitre 1 de cette thèse). La question du but justificatif général du mérite, de savoir pourquoi il est important de récompenser des individus selon leur mérite, nous occupera dans les chapitres 6 à 9 de cette thèse.

Ce chapitre sera consacré à la question de la distribution, c'est-à-dire à la question de savoir quels critères doivent être utilisés pour distribuer les biens. Il s'agit donc de la question de savoir comment reconnaître quels individus peuvent être considérés comme méritants. Bien sûr, ce problème est logiquement dépendant du traitement des buts justificatifs du mérite. Je pense cependant qu'il est facile de montrer que n'importe quelle théorie plausible du mérite doit reposer sur une pluralité de critères, et, en ce sens, il peut être utile de montrer quels critères devront être adoptés, quelle que soit la théorie justificatrice générale qui sera présentée dans les chapitres ultérieurs⁴⁷.

Je commencerai ici par un bref examen critique des deux critères les plus couramment adoptés pour l'identification des personnes méritantes, à savoir, l'effort et la contribution. Je montrerai qu'aucun de ces critères n'est véritablement plausible, parce qu'ils sont tous les deux soumis à des contre-exemples évidents : la récompense des efforts impliquerait de récompenser

⁴⁶ H. L. A. HART, *Punishment and Responsibility*, *op. cit.*, p. 4.

⁴⁷ On peut noter qu'il existe une grande confusion dans la littérature sur le mérite entre la question de la justification du mérite et la question de la distribution du mérite. Dans son article sur le mérite et les salaires, Owen McLeod présente trois grandes théories du mérite : le mérite comme effort, le mérite comme contribution, et le mérite comme compensation. O. MCLEOD, « Desert and Wages », *op. cit.* On voit bien qu'il s'agit d'une confusion : l'effort et la contribution sont bien des critères de distribution du mérite, mais ce n'est pas le cas de la compensation. La compensation est une justification du mérite : d'après cette théorie, il est bon de récompenser les individus méritants parce que cela sert à compenser les dommages qu'ils ont subis. En général, les théories du mérite comme compensation identifient deux critères principaux du mérite : le degré de sacrifice lié à une action, et les efforts subis par l'individu. Les efforts, comme critère du mérite, peuvent donc être justifiés par une théorie du mérite comme compensation.

fortement des personnes parfaitement inutiles, tandis que la récompense de la contribution pure conduirait à récompenser des individus ayant effectué une contribution importante par pur hasard.

Pour dépasser ces impasses, je proposerai un nouveau critère de mérite, à savoir l'identification du mérite à l'espérance de la contribution. Même s'il s'agit d'un critère apparenté aux critères traditionnels du mérite comme contribution, ce nouveau critère fait appel à des contrefactuels, et dépasse ainsi la dépendance que les théories du mérite comme contribution subissent vis-à-vis du hasard.

Dans un dernier temps, nous verrons cependant que le recours aux contrefactuels nécessaires pour estimer le mérite comme espérance de contribution rend difficile une évaluation exacte du mérite dans la vie courante. En conséquence, nous montrerons que les efforts et la contribution peuvent être utilisés comme approximations du mérite comme espérance de contribution.

Les critères traditionnels : l'effort et la contribution

Dans l'ensemble des critères traditionnels de mérite, deux critères ont eu une importance historique particulière : les efforts et la contribution. Une théorie du mérite uniquement basée sur le critère de l'effort a par exemple été défendue par Sadurski dans *Giving Desert its due*, ou par Ake dans « Justice as equality » ; une théorie de la contribution a été défendue par différents auteurs, le plus important étant David Miller dans *Market, State and Community* et dans *Principles of Social Justice*⁴⁸. Comme plusieurs auteurs l'ont fait remarquer, les théories « pures » de la contribution ou des efforts ne sont pas très convaincantes. Je vais rappeler les arguments les plus forts à l'encontre de chacune des théories pures, avant d'indiquer pourquoi une théorie mixte s'impose et fournit une alternative autrement plus satisfaisante.

Commençons par la théorie pure de l'effort. Selon cette théorie, les individus doivent être récompensés en proportion de leurs efforts effectués avec une intention positive en tête. Cette théorie est régulièrement justifiée en faisant appel à l'idée que les individus ne possèdent un parfait contrôle que sur les efforts qu'ils peuvent faire ; en conséquence, le seul critère possible du mérite consiste dans les efforts effectués.

⁴⁸ W. SADURSKI, *Giving Desert Its Due*, *op. cit.* ; C. AKE, « Justice as Equality », *op. cit.* ; David MILLER, *Market, state, and community: theoretical foundations of market socialism*, Oxford, Clarendon, 1990 ; D. MILLER, *Principles of Social Justice*, *op. cit.*

Quoique l'on pense de l'idée que le mérite doit uniquement prendre en compte ce qui est sous un parfait contrôle individuel, il faut tout de suite remarquer que la théorie du mérite comme effort livre des cas extrêmement contre-intuitifs. Certains contre-exemples sont même cités par les théoriciens du mérite fondé sur l'effort. Ainsi, Ake défend l'idée que si une personne dénuée de tout sens musical faisait autant d'efforts pour jouer du violon que Yehudi Menuhin, et si cette personne pensait vraiment qu'elle était en train d'offrir au monde une contribution importante, alors il serait juste de lui donner autant d'argent qu'à Yehudi Menuhin⁴⁹. Dit en d'autres termes, cette personne serait victime d'une injustice si elle ne gagnait pas un salaire considérable simplement en jouant mal d'un instrument.

Sans prendre un exemple aussi extrême, on peut imaginer deux employés d'une entreprise, qui font tous les deux beaucoup d'efforts, mais qui n'ont pas du tout le même niveau de compétences. L'un des deux est extrêmement compétent, et en conséquence apporte beaucoup de profits à l'entreprise. L'autre est très peu compétent, et, malgré ses efforts et sa bonne volonté, fait perdre beaucoup d'argent à l'entreprise. L'idée qu'il serait juste que les deux employés reçoivent le même salaire semble extrêmement contre-intuitif. De ce fait, toute théorie pure du mérite comme effort semble en rupture avec nos intuitions relatives au mérite.

Cependant, une théorie pure de la contribution ne s'en sort pas beaucoup mieux, et reste également opposée aux conceptions courantes du mérite. Le problème vis-à-vis de ces théories est qu'elles associent le mérite d'un individu avec ce que cet individu a effectivement produit. Or, tout travail productif semble ouvert à un certain nombre d'aléas ; même quelqu'un de compétent, faisant beaucoup d'efforts, peut par malchance aboutir à un résultat de faible qualité. Le monde de la recherche est évidemment un cas très net : un chercheur compétent peut s'aventurer dans de fausses pistes, en travaillant de nombreuses heures pour obtenir un résultat nul. A l'inverse, un chercheur peut être chanceux, et effectuer une grande découverte par pur hasard. Si l'on imagine deux chercheurs ayant effectué la même contribution, mais qui diffèrent dans les efforts qu'ils ont fait et dans leur niveau de compétence, il semble extrêmement contre-intuitif de dire qu'un chercheur incompetent, paresseux mais chanceux mérite la même récompense qu'un chercheur compétent et zélé.

Le problème posé par le hasard pour les théories pures de la contribution est renforcé par le fait que ces théories sont souvent des théories de la contribution entendue comme contribution marchande sur un marché libre et non régulé⁵⁰. Selon ces théories, le salaire mérité par un

⁴⁹ C. AKE, « Justice as Equality », *op. cit.*

⁵⁰ Là encore, le cas de la théorie du mérite proposée par David Miller représente le cas le plus net. D. MILLER, *Principles of Social Justice*, *op. cit.*

individu correspond à la valeur marginale de sa contribution économique. Cependant, même sur un marché idéal, la valeur de la contribution d'un travailleur est toujours produite par des facteurs complètement indépendants de sa volonté, voire même sans rapport avec ses talents, ses choix, et avec ce qu'il avait la capacité de prévoir. Si un individu travaille fortement pour produire un bien qui était en très forte demande, et si la demande pour ce bien chute subitement, alors la valeur marginale de son travail chutera complètement, alors même qu'il ne semble pas moins méritant pour autant.

Le fait que les idées courantes à propos de la justice rejettent cette réduction de la récompense du mérite à l'attribution du salaire marginal a été bien mise en avant par certaines expériences effectuées par Daniel Kahneman et ses collègues dans les années 1980. Dans un article de 1986, « Fairness as a Constraint on Profit Seeking », Kahneman et ses collègues ont posé à des personnes canadiennes recrutées par téléphone plusieurs questions visant à mettre en avant la rupture entre les conceptions courantes d'un juste revenu et les prédictions faites par une théorie basée sur l'idée d'un revenu marginal⁵¹.

La plus célèbre des questions posées par les auteurs porte sur le prix d'une pelle à neige après une tempête de neige : « Un commerce vend des pelles à neige pour 15\$. Le matin suivant une grande tempête de neige, le commerce augmente les prix jusqu'à 20\$. Notez cette action comme étant : Complément juste, acceptable, injuste, très injuste. » 82% des 107 personnes interrogées ont noté cette situation comme étant injuste. On peut supposer qu'ils considéraient que le commerçant ne méritait pas ce revenu supplémentaire. Pourtant, d'un point de vue strictement économique, il n'a fait que répondre à l'excès de demande.

Une autre question montre que les intuitions courantes n'acceptent pas non plus que des individus ou des entreprises profitent du manque de concurrence, sans pour autant proposer un produit de meilleure qualité. Voici une autre vignette proposée par les auteurs : « Une chaîne de primeurs possède des boutiques dans de nombreuses communautés. La plupart d'entre elles font face à une compétition de la part d'autres primeurs. Dans une communauté, la chaîne n'a pas de compétition. Même si les coûts et le volume de ventes est le même qu'ailleurs, la chaîne propose des prix qui sont en moyenne 5% plus chers que dans d'autres communautés. » 76% des 101 personnes interrogées ont considéré que cette décision est injuste. Là encore, il y a un écart entre ce qui serait prédit par une théorie de la contribution comme prix de marché et nos intuitions par rapport à ce que les commerçants méritent vraiment de recevoir.

⁵¹ Daniel KAHNEMAN, Jack L. KNETSCH et Richard THALER, « Fairness as a Constraint on Profit Seeking: Entitlements in the Market », *The American Economic Review*, 1986, vol. 76, n° 4, p. 728-741.

La contribution entendue en des termes purement économiques peut donc ne pas correspondre aux intuitions courantes relatives au mérite, parce qu'elle conduit le mérite d'un individu à être réduit à la merci du hasard. Il existe en effet un double hasard dans la contribution effectuée par l'individu. Du côté de l'offre, il existe un hasard dans la traduction des efforts et des capacités productives en biens effectivement produits. Du côté de la demande, il existe un hasard quant à la question de savoir si ce bien sera toujours populaire à l'avenir. Une théorie pure du mérite comme contribution, qu'il s'agisse d'une contribution marchande ou non marchande, dépend trop de la chance pour être véritablement plausible comme théorie du mérite.

Une nouvelle version du mérite comme contribution : le mérite comme espérance de contribution

Faisons le bilan de ce double examen. La théorie du mérite comme effort était inacceptable parce qu'elle ne prenait pas en compte le fait que le mérite doit avant tout récompenser une activité positive, et que tout effort dépensé, même avec une bonne intention, ne doit pas vraiment compter comme activité positive si l'individu en question est incompetent. La théorie du mérite comme contribution est inacceptable parce qu'elle ne prend pas en compte le fait qu'une contribution ne reflète que de façon très imparfaite le travail effectué par l'individu. Il s'agit donc de proposer des critères de mérite qui prennent en compte une activité vraiment positive, et qui soit vraiment sous le contrôle de l'individu.

Ces deux critères, liés à la positivité de l'action et au contrôle de l'individu, sont problématiques quand ils vont de pair. Il existe une double tension : d'une part, le mérite doit prendre en compte des critères internes propres à l'individu, à ses efforts et à sa capacité de contrôle de la situation. D'autre part, le mérite doit refléter un caractère extérieur, à savoir le côté positif d'une action ; or, le côté positif d'une action n'est jamais parfaitement prévisible à l'avance.

Pour résoudre cette tension entre l'idée de contrôle intérieur et d'utilité extérieure, je propose d'identifier le mérite à la productivité de l'action d'un individu, sans que cette productivité soit entendue dans un sens purement économique. Dans le sens où je l'entends, la productivité d'une action doit être comprise comme l'espérance de la contribution de l'action.

Avant de détailler plus avant ce qu'implique une telle proposition, il faut indiquer sa relation par rapport aux deux critères de la contribution et de l'effort. En un mot, ma proposition est un type de théorie de la contribution, mais qui tente de limiter la susceptibilité au hasard dont sont victimes les théories classiques de la contribution. Alors que les théories classiques de la contribution tendent généralement à identifier le mérite aux contributions effectives, sans prise en compte des aléas du hasard, ma proposition identifie le mérite aux contributions que le type d'action effectué tend à produire, indépendamment de savoir si ces actions ont dans la réalité produit une contribution positive ou non.

Reprenons maintenant les aspects techniques de la définition de la productivité d'une action. La productivité a d'abord été caractérisée comme « l'espérance de la contribution de l'action ». L'espérance est une notion probabiliste, qui correspond à l'idée d'une moyenne pondérée des différentes valeurs que peut prendre une variable aléatoire. L'espérance correspond grossièrement à la valeur moyenne que peut prendre une variable aléatoire, si l'on disposait de l'observation d'un très grand nombre d'apparitions de cette variable aléatoire.

Pour mieux faire comprendre cette idée, prenons l'exemple d'un individu qui effectue une activité productive. Il s'agit d'imaginer que l'action d'un individu peut avoir plusieurs conséquences. Pour simplifier la présentation, supposons que son action aboutit à une contribution pouvant prendre trois valeurs possibles : 0 (une action non productive, donc) si le travail échoue ; 1 (si le travail aboutit à un résultat positif, mais pas extraordinaire) ; et 10 enfin (si le travail aboutit à un résultat bien meilleur que prévu). Pour donner un ancrage à ces valeurs, on peut imaginer que 1 correspond à la valeur moyenne du travail effectué par des personnes dans ce secteur d'activité ; 10 correspondrait alors à une production 10 fois supérieure à la production moyenne. Supposons que l'action de l'individu a 20% de chances d'aboutir à un résultat de 0, 70% de chances d'aboutir à un résultat de 1, et enfin 10% de chances d'aboutir à un résultat de 10. Dans ce cas, l'espérance de contribution liée à ce travail est de : $0,2 \times 0 + 0,7 \times 1 + 0,1 \times 10 = 1,7$. Ces valeurs sont arbitraires, mais peuvent être modifiées sans changer l'argument fondamental. Ce type de valeur peut, de façon réaliste, se trouver dans le milieu de la recherche : la valeur 0 correspondrait à des recherches ayant abouti à un cul de sac ; la valeur 1 correspondrait à une découverte utile, mais qui ne change pas fondamentalement le domaine de recherche ; 10 correspondrait à une découverte importante.

Mobiliser l'idée d'espérance contributive permet de mettre en avant le fait qu'une action, ou un travail, amènent toujours à un résultat incertain ; cette idée permet d'établir une valeur de l'action qui ne dépend pas de ce que le hasard a effectivement mené à réalité.

Avant de continuer, trois choses sont à noter. Tout d'abord, il faut bien remarquer que la définition que je donne de la productivité d'une action implique qu'une action a été réalisée, ou est potentiellement réalisée. Il ne s'agit donc pas de prendre en compte simplement le fait qu'un individu est compétent, sans que ces compétences aient été mises en œuvre de façon concrète. Calculer l'espérance d'une action implique donc de considérer une action, et de prendre en compte la compétence de l'individu, ses efforts, des facteurs extérieurs, et de déterminer quelles sont les différentes contributions possibles associées à ces différents facteurs.

Deuxième point important, il faut également prendre en compte le fait que calculer l'espérance d'une action implique toujours d'associer des probabilités à des événements, comme par exemple la probabilité qu'un individu fasse un travail de valeur X. Le lecteur s'interrogera probablement sur la méthode à utiliser pour calculer ces probabilités. Une objection se présente effectivement immédiatement : si l'on imagine que nous habitons dans un univers déterministe, alors, de façon objective, il n'y a toujours qu'un seul événement qui peut arriver⁵². Il n'existe alors qu'un seul événement possible, qui est un événement certain, même s'il n'est pas possible de savoir à l'avance quel événement certain se réalisera. Si l'on défend cette idée, et qu'on considère qu'il n'existe qu'un seul événement possible possédant une probabilité 1, alors l'espérance de la contribution d'un individu se confondrait avec la contribution réalisée. La théorie du mérite comme productivité se réduirait à une théorie du mérite comme contribution.

La probabilité doit donc être déterminée de façon indépendante de l'événement qui s'est effectivement produit. Je propose donc de caractériser cette probabilité comme les croyances d'un observateur raisonnable, ayant une parfaite information relative à toutes les connaissances partagées par les personnes au moment de l'acte, et qui est capable de connaître parfaitement les compétences et les efforts de l'acteur. Il s'agit bien sûr, dans la plupart des cas, d'un observateur imaginaire, semblable au spectateur impartial de la théorie morale d'Adam Smith. Cette caractérisation de la probabilité permet de capturer les principaux éléments nécessaires à la construction d'une théorie plausible du mérite : tout d'abord, les différents éléments possibles sont évalués à partir du moment où l'acte est effectué, et non pas seulement après coup ; ensuite, cette évaluation des possibilités se fonde à la fois sur des éléments propres à l'individu, et sur des faits objectifs connus largement au moment de l'acte. Il n'est pas nécessaire d'avoir une connaissance parfaite de l'état de l'univers au moment de l'action pour pouvoir attribuer des

⁵² Bien sûr, le déterminisme n'est pas une hypothèse scientifique ayant le vent en poupe à l'heure actuelle. Cependant, le fait qu'un univers déterminisme pourrait réduire à néant la pertinence de la théorie défendue plus haut constitue en soi une menace suffisamment importante pour qu'il faille y répondre.

probabilités. En ce sens, une détermination des probabilités est exigeante (il faut avoir une connaissance approfondie de l'individu et du contexte de l'action) tout en restant possible (il n'est pas nécessaire d'avoir une connaissance divine).

Passons maintenant au troisième problème qu'il faut régler vis-à-vis de cette caractérisation du critère du mérite. Il faut bien remarquer que les critères définis jusque-ici ne règlent pas la question de savoir comment mesurer la « contribution » d'un individu. Comment mesurer la contribution d'un médecin, d'un enseignant, d'un ouvrier, d'un paysan, d'un cadre ? Je ne fournirai pas ici de solution à cette question délicate. Je me contenterai d'esquisser certains facteurs qui pourraient être pris en compte. Une théorie de la contribution devrait prendre en compte une liste de différents avantages, ou biens humains, qui peuvent être produits par des individus. Ces biens peuvent inclure, par exemple, la connaissance, la satisfaction des besoins, des données matérielles, ou la production de services ou de bien-être. Savoir comment mesurer la valeur de certains biens vis-à-vis d'autres bien doit nécessairement prendre en compte plusieurs facteurs, qui doivent inclure la demande relative à ces différents biens, les désirs que les individus ont que ces biens soient produits, et la facilité avec laquelle d'autres individus seraient prêts à produire ces biens.

La proposition de mesurer le mérite comme productivité pourra sembler extraordinairement complexe au lecteur. En vérité, elle l'est. Cependant, il est évident que cette complexité ne peut pas compter comme un argument décisif contre cette proposition. En effet, toute idée morale s'expose nécessairement à des complexités radicales dans leur mise en place. Prenons l'exemple de l'idée d'égalité, dont les complexités ont été magistralement mises en évidence par Larry Temkin dans son livre *Inequality*⁵³. Mesurer le degré d'inégalités au sein d'une société présuppose de comparer différents types de bien possédés par des individus (comme les loisirs, l'argent, le prestige), de prendre en compte différents types de désavantages (la pénibilité au travail, de potentiels handicaps physiques ou mentaux), et de mélanger tous ces facteurs pour faire l'estimation des biens et des désavantages au sein de la société. Ensuite, une fois qu'une telle estimation générale des biens et des désavantages a été faite, il s'agit de déterminer quels sont les sujets de l'égalité ; est-ce que l'égalité est un attribut des individus ou des groupes ? Faut-il mesurer l'inégalité comme écart par rapport à la moyenne, ou comme écart par rapport à l'individu le plus riche d'une société ? Est-ce qu'il est préférable d'avoir un individu au sommet de la société, ou plusieurs ? L'estimation de l'inégalité pose une multitude de questions, sans qu'il soit évident que nous disposions d'idées fermes sur ce sujet. Même les

⁵³ Larry S. TEMKIN, *Inequality*, New York, Oxford Univ. Press, 1996.

idées les plus courantes et les plus intuitives révèlent une complexité fondamentale dès qu'on les examine ; il n'est pas surprenant que le mérite recèle les mêmes difficultés.

Malgré les complexités qui viennent d'être mises en avant, la proposition de comprendre le mérite comme productivité d'une action est extrêmement basique. Fondamentalement, il s'agit d'une distinction entre productivité sur le long terme, ou tendance à créer des actions positives sur le long terme, et contribution effective ; cette distinction est incroyablement simple. Elle a toutefois été négligée dans les débats sur le mérite comme contribution. La plupart des définitions du mérite comme contribution tendent à se concentrer sur ce qui a été effectivement produit par l'individu, et non sur ses contributions potentielles. En témoignent les différentes caractérisations du mérite comme contributions recensées par McLeod dans son article « Desert and Wages », tirées de trois auteurs différents :

La récompense d'un individu devrait dépendre de la valeur de la contribution effectuée en faveur du bien-être social lors de son activité de travail.

La justice, d'après ce principe [de contribution], exige que chaque travaille obtienne exactement la proportion de la richesse nationale qu'il a lui-même créé.

D'après la théorie de la contribution, le salaire que vous méritez est une fonction de la quantité de contribution que vous avez effectuée⁵⁴.

Dans ces trois cas, la contribution auxquels les auteurs font références est toujours une contribution qui a été effectivement produite. Aucune de ces définitions ne tient compte de ce que les actions auraient eu tendance à produire sur le long terme. De ce fait, toutes ces théories s'exposent aux reproches d'une trop grande exposition au hasard. Ces théories sont donc conduites à défendre des exemples parfaitement contre-intuitifs, qui ont été exposé plus haut.

A l'inverse, la théorie du mérite comme productivité présentée ici n'est victime d'aucun des contre-exemples. Nous avons vu que le mérite comme effort était victime du fait que cette théorie exige que des travailleurs sans aucune productivité reçoivent le même salaire que des travailleurs très efficaces. La théorie du mérite comme productivité évite évidemment ces cas. De plus, elle évite également les problèmes auxquels s'exposent la théorie du mérite comme contribution effective. Nous avons vu que la théorie du mérite comme contribution effective était problématique, parce qu'elle défend l'idée que la contribution a tout autant de valeur si

⁵⁴ O. MCLEOD, « Desert and Wages », *op. cit.*, p.208-209. La première caractérisation du mérite comme contribution est de David Miller, la deuxième de Joel Feinberg, la troisième de Karol Soltan. La définition proposée par McLeod dans son article fait aussi référence à la contribution effectivement produite par les individus, et non pas à l'espérance de la contribution. David MILLER, *Social Justice*, Oxford, Oxford University Press, 1979 ; Joel FEINBERG, *Social philosophy*, Englewood Cliffs, NJ, Prentice Hall, 1973 ; Karol Edward SOLTAN, *The causal theory of justice*, Berkeley, Univ. of California Press, 1987.

elle est le résultat du pur hasard, que si elle était le résultat d'un travail fondé sur l'effort et la compétence. Parce que la théorie du mérite comme productivité se concentre sur les contributions probables au moment où un individu a effectué une action, elle ne s'expose pas aux mêmes difficultés. En effet, un individu compétent et effectuant des efforts aura en général une bien plus grande probabilité d'apporter une contribution positive qu'un individu sans compétence et ne faisant pas d'efforts. En ce sens, la théorie du mérite comme productivité ne s'expose pas à une surpondération de l'importance du hasard dans la vie morale.

La théorie du mérite comme productivité sert à résoudre les conflits générés par les théories rivales. Cependant, il faut bien noter qu'elle laisse planer une très forte incertitude dans la mesure du mérite. Étant donné que je considère qu'une incertitude dans la mesure du mérite est inévitable, je ne pense pas qu'il s'agisse d'un problème décisif. Il faut cependant noter que l'incertitude dans la mesure de la productivité des actions individuelles permet de défendre l'importance de se servir d'heuristiques dans la mesure de la productivité. En conséquence, on peut réhabiliter les critères d'effort et de contribution, mais seulement comme approximations de la théorie authentiquement valable. La section suivante va détailler davantage les difficultés pratiques auxquelles s'expose une théorie du mérite comme productivité.

L'incertitude de la mesure du mérite

Nous avons vu que le critère du mérite comme productivité s'expose à un certain nombre de difficultés. J'ai indiqué comment je pense que ces difficultés pourraient se résoudre, et j'ai montré pourquoi ce nouveau critère de mérite comme productivité est supérieur aux critères précédemment utilisés de mérite comme effort et de mérite comme contribution. J'ai réservé à part une autre difficulté, qui correspond au fait que ce nouveau critère rend difficile l'identification du mérite d'un individu dans un cas précis, et implique qu'il faut se servir d'heuristiques, ou de critères approximatifs, pour mesurer le mérite d'un individu.

Si l'on identifie le mérite d'une action à ce qu'une action aurait dû produire, étant donné les capacités d'un individu, ses efforts, et les circonstances extérieures telles qu'elles étaient accessibles à un individu informé de la situation, alors on voit bien que l'on rentre dans une zone difficile, où il est presque impossible de déterminer ce qu'un individu mérite effectivement. En effet, une situation ne se passe presque jamais telle qu'elle aurait dû se passer, dans le sens où il entre toujours dans l'action des hasards et des incertitudes. En termes

probabilistes, la valeur effective d'une action, considérée comme une variable aléatoire, n'est jamais exactement égale à son espérance mathématique ; la contribution que fera un individu sera toujours en dessous ou au-dessus de ce qu'elle est en moyenne. Un professeur enseignant à ses élèves réussira toujours plus ou moins ses cours ; certains cours seront très réussis, d'autres beaucoup moins, sans qu'il soit toujours prévisible à l'avance de savoir à quel point un jour de classe en particulier sera un succès. L'apprentissage effectif des élèves dépend de multiples facteurs ; même si les capacités générales et les efforts du professeur ont une importance décisive, de nombreux accidents extérieurs interviennent également.

Pour toutes ces raisons, identifier le niveau de mérite d'une action est extrêmement difficile, si l'on suit les critères ici proposés. On voit alors qu'il est nécessaire de pouvoir mobiliser des critères simples pour juger de la productivité d'une activité. Ces critères simples permettraient de juger d'un cas particulier même si l'on ne dispose pas de tous les éléments qui seraient nécessaires pour effectuer un jugement parfaitement informé. Il s'agit donc de trouver des éléments approximatifs, qui soient suffisamment corrélés avec le mérite authentique d'une action pour que l'on puisse s'en servir dans la vie ordinaire. Il s'agit donc de trouver des critères « heuristiques », qui évitent d'effectuer les recherches intensives liées à un choix parfaitement informé et rationnel.

Quels sont ces critères approximatifs du mérite, qui peuvent servir à identifier la productivité d'une action dans des contextes quotidiens ? Plusieurs exemples viennent à l'esprit, et ils nous ramènent vers les critères traditionnels du mérite : l'effort et la contribution. En général, il existe naturellement une forte corrélation entre les efforts qu'un individu fait et la contribution que ses actions tendent à produire. De façon similaire, même si la contribution finale d'un individu ne correspondra jamais exactement à la valeur moyenne de son action, il existera une très forte corrélation entre productivité et contribution effective. Les deux critères traditionnels possèdent donc une utilité, mais pas en tant que critères fondamentaux ; ils ne possèdent de valeur qu'en tant qu'ils servent d'approximation à un critère plus important, celui de la productivité.

Voir dans la contribution une approximation de la productivité nous renvoie vers des problèmes philosophiques classiques. En effet, on retrouve ici les mêmes difficultés étudiées dans les situations de fortune morale. L'idée de fortune morale renvoie à l'idée que le hasard a un impact fort sur nos évaluations morales, et qu'il est difficile de se détacher de l'influence du résultat final sur nos évaluations, quand bien même nous avons tendance à considérer que le hasard ne devrait pas avoir d'influence sur l'évaluation de la moralité d'un comportement. L'exemple classique consiste à imaginer deux personnes rentrant chez elles après une soirée très arrosée. Sur la route, ils s'endorment tous les deux au volant. Après s'être réveillé, l'un

d'entre eux se rend compte qu'il a roulé dans les champs, sans provoquer de gros dégâts ; l'autre se rend compte qu'il a écrasé une petite fille. Alors même que les deux individus ont effectué exactement les mêmes actions, nos sentiments de condamnation sont beaucoup plus importants pour l'un que pour l'autre. Le hasard semble avoir un impact important sur nos condamnations morales⁵⁵.

Le même problème se pose évidemment dans le cas général du mérite, si l'on considère que la production effective est un critère majeur comme approximation de la productivité. On aura tendance à juger qu'un individu qui aura fait une importante contribution sera plus méritant qu'un individu qui aura fait une contribution moyenne, même dans les cas où les deux individus ont fait les mêmes efforts ou ont le même degré de compétence. Comment expliquer ces cas ? Est-ce qu'ils ne constituent pas une objection majeure à la conception du mérite comme productivité ?

Pour répondre à ces questions, et donc pour expliquer pourquoi le résultat final a une telle importance sur nos évaluations morales, il est utile d'examiner l'analyse qu'Adam Smith a fait du problème de la fortune morale dans sa *Théorie des sentiments moraux*. Adam Smith cherche à expliquer pourquoi, dans le cas des jugements moraux, nous devons faire appel aux résultats, et non simplement aux intentions. Dans notre cas, il s'agirait d'expliquer pourquoi nous devons nous servir de la contribution effective comme approximation de la productivité, et non pas simplement des efforts et des intentions déclarées des participants.

Pour résoudre le problème posé par les cas de fortune morale, Smith fait appel à trois raisons pour lesquelles nous nous attachons autant aux actes, et non simplement aux pures intentions⁵⁶. La première raison présentée par Adam Smith fait appel au fait que les intentions des individus sont cachées. Il est beaucoup plus difficile de savoir ce qu'un individu avait en tête que de savoir ce qu'il a fait. Dans le cas des efforts, on peut aussi considérer qu'il est plus facile de mesurer les contributions d'un individu que ses efforts. Il peut être dur de savoir si quelqu'un a fait beaucoup d'efforts, ou très peu, si l'on ne l'a pas observé de façon extrêmement précise pendant une très longue période de temps. Autrement dit, il est souvent plus facile de juger quelqu'un par sa contribution que par ses efforts ou par ses intentions. En dehors du critère de facilité, Adam Smith souligne les dangers présentés par une théorie qui ne jugerait les individus que sur leurs intentions ou sur leurs efforts : ceux-ci étant largement cachés, il faudrait non

⁵⁵ Le lecteur intéressé pourra se référer aux articles classiques de Nagel et Williams sur le sujet. B. A. O. WILLIAMS et T. NAGEL, « Moral Luck », *Proceedings of the Aristotelian Society, Supplementary Volumes*, 1976, vol. 50, p. 115-151.

⁵⁶ A. SMITH, *Théorie des sentiments moraux*, op. cit., p. 155-156.

seulement s'exposer à une grande incertitude, mais également exercer un contrôle inquisiteur sur les actions individuelles pour réussir à découvrir leurs véritables efforts et intentions. En ce sens, une évaluation basée sur la contribution respecte davantage l'autonomie et la liberté des individus.

La deuxième raison avancée par Adam Smith concerne le fait que, selon ses propres mots, « l'homme a été fait pour l'action ». Autrement dit, la récompense des pratiques de mérite ne prend sens que dans un monde où il est possible de bien agir, et de contribuer à l'amélioration du bien-être d'autres individus. En ce sens, il est important d'instituer des pratiques de récompenses qui soulignent qu'il n'est pas suffisant de se contenter de bonne volonté, mais qu'il est important de rechercher le plus possible la réalisation effective d'un progrès. Si l'on interprète cette raison en termes d'approximation du mérite comme productivité, on peut dire que le fait que, si un individu plein de bonnes intentions et plein d'efforts ne produit rien sur le long terme, alors il s'agit d'un signe fort que cet individu était en réalité un vertueux inutile, dont les actions possédaient une productivité nulle. En ce sens également, il est préférable d'attribuer une grande importance à la production, et non aux seules intentions.

En plus de ces deux raisons, Smith fait appel à une troisième idée pour expliquer pourquoi nous attribuons une telle importance aux actes, et non seulement aux conséquences. Comme cette troisième raison s'applique uniquement aux cas d'accidents et de dommages (comme les cas classiques de fortune morale), elle nous servira moins pour la constitution d'une théorie générale du mérite ; elle reste, cependant, d'une certaine utilité. D'après Smith, le fait que nous nous sentons coupables des maux causés involontairement à quelqu'un nous conduit à être très circonspect vis-à-vis du bonheur d'autrui. Selon Smith, comme nous avons l'impression d'avoir commis un sacrilège si jamais nous attendons au bien-être d'autrui, même par accident, cela nous conduit généralement à être très prudent dans nos actions si jamais celles-ci peuvent avoir des conséquences négatives pour autrui.

En faisant appel à cette troisième raison, Adam Smith avait manifestement en tête des cas de dommages. Il est cependant possible de tirer une leçon générale des arguments présentés. Récompenser la contribution permet de rendre un retour communicatif très simple et très clair : le bien-être des individus est valorisé, et tout dommage est un sacrilège. En cela, récompenser le mérite possède un avantage psychologique, en tant qu'il permet de tracer une ligne claire entre les bonnes actions, qui servent le bien-être d'autrui, et les mauvaises actions, qui le desservent. En ce sens, la récompense des contributions possède une simplicité psychologique inégalée.

En d'autres termes, l'importance des contributions renvoie à trois faits fondamentaux : tout d'abord, les contributions renvoient à un fait facilement observable, qui nous dispense d'exercer une attitude inquisitoriale vis-à-vis des actions d'autrui. Deuxièmement, la mesure de la contribution rend saillant le fait que la récompense du mérite prend sens dans la mesure où des actions peuvent améliorer le bien-être d'autres personnes ; la récompense de la contribution rend manifeste le but ultime de tout système de récompense. Enfin, troisièmement, se fier aux contributions présente une grande simplicité en termes de retour communicatif : si quelqu'un a été récompensé, c'est pour sa contribution, et cela rend saillant le fait que défendre le bien-être d'autrui possède de la valeur. Prises ensemble, ces trois raisons justifient l'importance que l'on accorde généralement à la contribution individuelle, par-delà les simples efforts et intentions.

Toutes ces raisons expliquent donc pourquoi la contribution constitue généralement une approximation satisfaisante de la productivité d'une action. Cependant, ces facteurs ne veulent certainement pas dire qu'il n'est jamais bon de prendre en compte les efforts ou l'intention d'un individu. En effet, les objections formulées plus haut à l'encontre de la production comme critère unique du mérite valent toujours. Se concentrer uniquement sur la production d'un individu risquerait de nous exposer à accorder une importance démesurée au hasard.

De fait, comme nous venons de le voir, le hasard a toujours un rôle à jouer dans les jugements moraux ; la contribution est toujours modelée par le hasard, et le simple fait de prendre la contribution comme approximation de la productivité implique que nous nous exposons toujours au risque de la fortune morale. Mais nous pouvons chercher à atténuer le risque de la fortune morale. Nous pouvons nous servir de plusieurs critères qui sont des signes de la productivité, et en les pondérant nous pouvons ainsi obtenir une meilleure approximation que si nous ne nous étions préoccupés que d'un seul moyen d'évaluer la productivité. En ce sens, il est utile d'effectuer un jugement tenant aussi bien en compte les compétences, les intentions, les efforts, et la production finale, afin de chercher à trouver le jugement le plus précis de la productivité réelle d'un individu.

Prenons un exemple simple. Un employé d'une entreprise vient d'arriver ; il n'y a absolument aucune information sur lui. Il effectue un travail en solitaire ; ce travail est de très mauvaise qualité. L'employé affirme avoir fait beaucoup d'efforts, mais que la tâche était beaucoup plus complexe que prévu. Il est difficile de juger de ce cas ; il n'y a pas assez d'information pour juger véritablement si l'employé a fait un travail productif ou non. Mais si l'on change l'histoire, et que l'on précise que l'employé avait fait preuve à de nombreuses reprises de sa productivité, de sa compétence et de ses efforts, alors le jugement change. Il devient très fortement probable que l'employé s'est réellement attaché à un travail important,

mais qui, par malchance, s'est avéré non productif. Il semblerait alors injuste de lui tenir rigueur d'avoir fait un travail qui, en général, est très productif. Dans ce cas, il semble légitime de ne pas tenir compte du résultat médiocre, et de dire que l'employé mérite d'être bien payé, même s'il n'a rien produit par malchance⁵⁷.

Cet exemple montre donc qu'une pondération de différents facteurs doit permettre de mesurer la productivité d'un individu. Une mesure approximative du mérite n'est donc en rien impossible. Il n'en reste pas moins qu'adopter un critère réaliste du mérite ne peut éviter des cas contre-intuitifs, liés à la place du hasard dans la vie morale. Comme Saint Augustin l'avait démontré de façon plausible dans *la Cité de Dieu*, adopter une théorie morale pour le monde terrestre conduit nécessairement à nous exposer régulièrement à des injustices⁵⁸. Cependant, sauf à remettre la réalisation de la justice à Dieu, il semble nécessaire d'accepter des approximations acceptables de critères légitimes du mérite.

Conclusion

Ce chapitre a fourni une brève discussion des critères susceptibles de déterminer la distribution des récompenses du mérite. Dans un premier temps, en reprenant les concepts mobilisés par Hart dans ses prolégomènes à une théorie de la peine, j'ai opéré une distinction entre la justification du mérite et les critères permettant de l'identifier. J'ai ensuite procédé à une critique des deux critères traditionnels, l'effort et la contribution, ce qui a permis de démontrer qu'aucun des deux critères n'est vraisemblable en tant que critère unique, puisque les efforts sont compatibles avec le fait de récompenser des actions parfaitement inutiles, et puisque la contribution est soumise à l'effet du hasard. Pour dépasser ces limites, j'ai proposé, dans un deuxième temps, d'identifier le critère du mérite à la productivité, entendue comme espérance de contribution produite par les actes individuels. Ce nouveau critère présente l'avantage d'échapper aux contre-exemples auxquels étaient soumis les critères traditionnels de l'effort et de la contribution. Cependant, ces deux derniers critères présentaient un avantage net vis-à-vis du critère de la productivité : ils étaient beaucoup plus facilement identifiables,

⁵⁷ Et à l'inverse, naturellement, si cet employé avait eu une histoire d'inefficacité et d'incompétence, le jugement sur le travail inutile changerait du tout au tout. Il s'agirait d'une nouvelle preuve que cet employé ne s'adonne à aucun travail productif, et qu'il ne mérite pas un salaire remarquable.

⁵⁸ AUGUSTINUS, *La cité de Dieu*, Paris, Gallimard, 2008.

puisque'ils ne dépendaient pas de l'identification d'événements contrefactuels. En conséquence, la dernière partie de ce chapitre a consisté à mettre en avant le fait que la contribution et les efforts pouvaient servir de critères approximatifs et heuristiques pour déterminer le mérite réel d'un individu.

Ce chapitre a cherché à déterminer quels critères du mérite correspondent à nos intuitions ordinaires. Conformément à la distinction entre critère et justification du mérite, je n'ai pas véritablement tenté de démontrer que récompenser des individus en fonction de ces critères possède une véritable valeur. La distinction entre récompense des efforts et récompense de la contribution a concentré une majeure partie des débats philosophiques autour du mérite. Pour cette raison, j'ai pensé qu'il était important de fixer l'attention du lecteur sur les critères les plus plausibles, afin qu'il ait une idée précise de ce qui était entendu ici par l'idée de jugement de mérite. Cependant, dans une perspective philosophique, l'identification des critères est secondaire par rapport à la question de la justification générale du mérite. S'il était démontré que le mérite était sans valeur, alors l'identification de critères de mérite demeurerait sans grand intérêt. Le lecteur peut cependant se rassurer : les prochains chapitres seront consacrés à l'élaboration d'une justification générale du mérite, qui démontrera que les critères qui ont été ici mis en avant sont indispensables pour la vie morale ordinaire.

Chapitre 3

La centralité du mérite dans les conceptions morales ordinaires

Dans le premier chapitre, j'ai réduit progressivement l'étendue du concept de mérite, afin d'obtenir un concept opérationnel qui soit à même de caractériser uniquement des jugements portant sur le caractère approprié d'une récompense ou d'une peine dépendante du caractère positif ou négatif d'une action, et satisfaisant un système d'ordonnement minimal. Avec cette conceptualisation, il devient clair que le mérite ne se confond pas avec l'ensemble des jugements moraux ou des jugements de justice, mais renvoie à un sous-ensemble bien particulier. Le lecteur est alors en droit de se demander dans quelle mesure cette conception bien délimitée du mérite correspond à un type de jugement important. Il s'agit alors de déterminer dans quelle mesure le mérite représente un aspect fondamental des conceptions morales ordinaires.

Ce chapitre vise à présenter une partie des preuves empiriques démontrant le fait que le mérite est un principe de justice extrêmement populaire et répandu. Pour juger le caractère répandu d'une idée, on peut adopter différents niveaux d'analyse. Je présenterai ici un ensemble de preuves relatives aux jugements explicites des individus A) dans les pays occidentaux, puis B) dans des pays non-occidentaux.

Ce chapitre fournira également l'occasion d'une brève introduction à l'étude de la perception de la justice sociale dans son ensemble. En 1992, David Miller a publié un article influent où il défendait l'idée que les études de psychologie sociale ont montré que les individus n'adoptent pas la défense d'un principe unique de justice, mais sont favorables à une pluralité de principes de justice⁵⁹. Plus précisément, Miller a défendu l'idée que les personnes adhèrent en général à trois principes de justice différents : l'égalité, les besoins, et le mérite. Des recherches récentes semblent plutôt confirmer le constat, même si d'autres chercheurs ont défendu l'idée que

⁵⁹ David MILLER, « Distributive Justice: What the People Think », *Ethics*, 1992, vol. 102, n° 3, p. 555-593.

l'égalité n'était en rien une idée populaire⁶⁰. Il s'agit ici de mettre en avant l'idée qu'une considération de l'importance du mérite doit toujours être relativisée par la prise en compte d'autres principes de justice.

Même si les données réunies dans ce chapitre convergent pour démontrer que le mérite est central dans les conceptions ordinaires dans l'ensemble des sociétés étudiées jusqu'à présent, il faut noter dès à présent un certain nombre de limites. Tout d'abord, les études en psychologie des jugements de mérite ont porté de façon beaucoup plus étendue sur les pays occidentaux que sur les pays non-occidentaux ; les données restent globalement pauvres dans ce dernier cas. Ensuite, les données obtenues dans des projets interculturels tendent à montrer un impact net de la culture, qui laissent entendre que le mérite peut posséder une importance moindre dans certaines sociétés non-occidentales, comparées aux États-Unis. Cependant, je procéderai dans ce chapitre à une nouvelle analyse d'enquêtes internationales, et je montrerai que, même s'il existe de vraies différences culturelles dans la perception de ce qui constitue un salaire juste, ces différences laissent transparaître une très forte convergence des opinions à travers les différents pays⁶¹.

La centralité du mérite dans les conceptions ordinaires : les pays occidentaux

Au cours des années 60 – 70 a fleuri un courant de recherche en psychologie portant sur les perceptions de justice dans le cas d'une production limitée à deux individus ou à un groupe restreint de personnes. Il s'agissait donc de l'étude de l'usage de principes de justice à un niveau

⁶⁰ Contre l'idée que l'égalité est un principe de justice populaire, le lecteur pourra consulter les articles suivants : James KONOW, « Which Is the Fairest One of All? A Positive Analysis of Justice Theories », *Journal of Economic Literature*, 2003, vol. 41, n° 4, p. 1188-1239 ; Christina STARMANS, Mark SHESKIN et Paul BLOOM, « Why people prefer unequal societies », *Nature Human Behaviour*, 2017, vol. 1, n° 4, p. 0082.

⁶¹ Un phénomène sur lequel je ne reviendrai pas dans ce chapitre, mais qui témoigne bien de l'importance de l'idée de mérite, est ce qui a été appelé « croyance en un monde juste ». Originellement étudié par Melvin Lerner dans les années 1970, le phénomène de croyance en un monde juste correspond à la tendance à évaluer négativement les victimes. Lerner avait fait l'hypothèse que les participants à ses expériences dévalorisaient inconsciemment les victimes afin de pouvoir croire que celles-ci avaient mérité ce qui leur arrivait. L'idée de croyance en un monde juste a depuis fait l'objet d'études intensives. Pour une présentation en Français des recherches, voir M. DURU-BELLAT, *Le mérite contre la justice*, op. cit. ; Melvin J. LERNER, *The belief in a just world: a fundamental delusion*, New York, NY, Plenum Press, 1980 ; Melvin J. LERNER, « The Two Forms of Belief in a Just World », in Leo MONTADA et Melvin J. LERNER (dir.), *Responses to Victimization and Belief in a Just World*, Boston, MA, Springer US, 1998, p. 247-269 ; Carolyn L. HAFFER et Laurent BÈGUE, « Experimental Research on Just-World Theory: Problems, Developments, and Future Challenges. », *Psychological Bulletin*, 2005, vol. 131, n° 1, p. 128-167.

local. A partir des années 80-90, un nouveau courant de recherche a tenté d'étudier les intuitions regardant la justice au sein de la société dans son ensemble. Je présenterai brièvement ces champs de recherche tour à tour. Dans les deux cas, des preuves convergentes ont montré l'importance du mérite dans le public occidental.

Ces recherches ont visé à évaluer l'importance des idées de justice, d'égalité et de besoins dans les attributions de la légitimité des récompenses attribuées à divers individus. Ces expériences ont généralement fait usage d'histoires courtes. On peut diviser ces histoires en deux grands ensembles : dans un premier type d'expériences, les sujets devaient attribuer une note à des individus qui avaient effectué une certaine tâche, généralement dans un contexte sportif ou scolaire. Dans ce premier ensemble d'expériences, il n'y avait donc pas de bien à distribuer, et l'attribution des notes n'était pas coûteux. Un second ensemble d'expériences s'est servi d'un paradigme de travail, décrivant un groupe de deux ou plusieurs individus produisant un bien en commun. Par exemple, dans plusieurs expériences, la vignette décrit deux individus écrivant un livre en commun, et le participant est chargé d'évaluer la justice de la distribution des bénéfices réalisés⁶².

Dans le cas de ces deux paradigmes expérimentaux, les résultats sont largement concordants. En fonction de la situation, les participants prennent en compte à la fois les efforts et la contribution de chaque individu. En général, un individu faisant plus d'efforts sera davantage récompensé, ou obtiendra une meilleure note, qu'un autre individu ayant fait moins d'effort ; de la même manière, quelqu'un qui aura davantage contribué à la qualité d'un livre sera davantage rémunéré, même s'il a effectué les mêmes efforts et avait simplement de plus grandes capacités. Un tel résultat fondamental a été observé à travers tous les types de situations décrites, qu'il s'agisse de compétitions sportives, d'examens scolaires, ou de contributions dans un contexte de travail. Même si un tel résultat est clair et sans appel, trois points doivent être précisés, qui délimitent la portée d'une telle conclusion en faveur de la popularité de conceptions méritocratiques.

⁶² Au sein d'une littérature très vaste, le lecteur pourra consulter Gerald S. LEVENTHAL et James W. MICHAELS, « Locus of cause and equity motivation as determinants of reward allocation », *Journal of Personality and Social Psychology*, 1971, vol. 17, n° 3, p. 229-235 ; Stanley REST, Richard NIERENBERG, Bernard WEINER et Heinz HECKHAUSEN, « Further evidence concerning the effects of perceptions of effort and ability on achievement evaluation. », *Journal of Personality and Social Psychology*, 1973, vol. 28, n° 2, p. 187-191 ; Egon KAYSER et Helmut LAMM, « Input integration and input weighting in decisions on allocations of gains and losses », *European Journal of Social Psychology*, 1980, vol. 10, n° 1, p. 1-15 ; Helmut LAMM, Egon KAYSER et Volker SCHANZ, « An Attributional Analysis of Interpersonal Justice: Ability and Effort as Inputs in the Allocation of Gain and Loss », *The Journal of Social Psychology*, 1983, vol. 119, n° 2, p. 269-281 ; W. I. GRIFFITH, « The effect of inputs and performance outcomes on allocation decisions », *Social Justice Research*, 1990, vol. 4, n° 1, p. 35-47.

Tout d'abord, il faut remarquer que, même si les personnes ayant apporté une plus grande contribution ou ayant fait plus d'efforts sont mieux rémunérées que d'autres, cette rémunération supérieure n'est pas strictement proportionnelle à la contribution fournie ou aux efforts investis. Un résultat constant de ces études est que la distribution jugée la plus juste se situe toujours entre une distribution égale et une distribution proportionnelle aux efforts ou à la contribution. Dans une expérience relativement typique, effectuée en 1990 auprès de 300 étudiants américains, Griffith a présenté une vignette décrivant deux individus travaillant en commun et devant se partager 80 dollars⁶³. L'un des individus était toujours décrit comme ayant effectué plus d'efforts ou ayant fait une contribution plus forte que l'autre, au point d'avoir contribué aux trois quarts de la production. Dans ce cadre, une distribution proportionnelle consisterait à lui attribuer 60 des 80 dollars. Cependant, la part moyenne qui lui était attribuée a varié entre 49 et 58 dollars, en fonction des différentes conditions expérimentales (qui manipulaient notamment le degré d'amitié entre les deux individus). Il est possible d'interpréter de tels jugements comme offrant un compromis entre l'application des principes de justice distincts que sont l'égalité et le mérite. Plus généralement, d'autres études, qui manipulaient le degré de besoins ressentis par différents individus, ont également montré que les participants attribuent davantage d'argent à un travailleur si ce dernier est plus dans le besoin qu'un autre travailleur. C'est sur la base de telles observations que Miller a construit sa théorie d'un pluralisme des opinions ordinaires relatives à la justice entre adhésion au besoin, au mérite et à l'égalité.

Deuxièmement, il faut noter qu'il existe de fortes différences entre les réactions adoptées par divers participants. Il existe toujours une portion des participants qui préfèrent et considèrent comme plus juste une distribution égale plutôt qu'une distribution inégale des revenus. Par exemple, dans un article de 1983 portant sur des étudiants allemands, Lamm, Kayser et Schanz ont utilisé une vignette décrivant deux personnes ayant vendu un livre écrit en commun, et ayant réalisé un certain type de bénéfices⁶⁴. Même dans le cas où un individu a fait plus d'effort qu'un autre (travaillant 16 jours au lieu de 8), 16 participants sur 24 ont ainsi jugé qu'une distribution égale était la plus juste, tandis que les 5 autres participants jugeaient qu'une distribution parfaitement proportionnelle, accordant les deux tiers des bénéfices à l'individu ayant fait plus d'efforts, était la plus juste (les 3 autres participants ont adopté une position intermédiaire). Des études de ce type peuvent donc observer une division assez nette des participants entre différentes positions, même si la proportion d'individus adoptant une position égalitaire varie en fonction des expériences.

⁶³ W. I. GRIFFITH, « The effect of inputs and performance outcomes on allocation decisions », *op. cit.*

⁶⁴ H. LAMM, E. KAYSER et V. SCHANZ, « An Attributional Analysis of Interpersonal Justice », *op. cit.*

Cependant, il faut tout de suite remarquer que la division des opinions n'est pas si extrême qu'on pourrait le penser à première vue. Certaines études ont ainsi bien mis en avant qu'il existe un relatif consensus entre personnes situées à la droite et à la gauche de l'échiquier politique vis-à-vis de la question du mérite. Ainsi, Lisa Farwell et Bernard Weiner, dans un article publié en 2000⁶⁵, ont demandé à 96 étudiants américains d'imaginer qu'ils étaient membres d'une association caritative, et qu'ils devaient distribuer de l'argent à deux personnes au chômage⁶⁶. Les deux personnes décrites variaient dans leur responsabilité vis-à-vis de leur condition : l'un des deux chômeurs avait été licencié parce qu'il faisait peu d'efforts, tandis que l'autre avait perdu son emploi parce que son entreprise avait fait faillite. Les auteurs ont alors demandé aux différents participants d'indiquer comment ils distribueraient l'argent entre les deux, comment un conservateur typique le distribuerait, et comment un libéral typique (dans le contexte américain, une personne de gauche) le ferait. Dans tous les cas, les participants devaient indiquer combien d'argent ils devraient donner au chômeur sur une échelle allant de 1 (le moins d'argent possible) à 11 (autant d'argent que possible). Ils ont ensuite demandé aux participants leur affiliation politique.

Les résultats de cette étude indiquent que les différences entre conservateurs et libéraux sur cette question sont bien moindres qu'attendu. Dans le cas du chômeur qui n'est pas responsable de son sort, l'étude ne révèle aucune différence entre libéraux et conservateurs, la moyenne étant dans les deux cas à 9, indiquant une forte volonté d'aider la personne. Dans le cas du chômeur qui avait été licencié pour manque d'effort, l'opinion politique conduit bien à une différence, les libéraux donnant un score moyen de 5,2 et les conservateurs de 3,9. Cependant, deux points sont à noter. Tout d'abord, il existe un consensus au travers du spectre politique sur le fait qu'une aide est beaucoup moins légitime dans le cas de la personne responsable que dans le cas de la personne non responsable de la perte d'emploi. Ensuite, la différence entre conservateurs et libéraux est beaucoup plus faible que celle qui était attendue par les participants de l'enquête. Libéraux comme conservateurs s'attendaient à ce qu'un conservateur typique donne un score autour de 2 ou 2,5 dans le cas du chômeur responsable de son licenciement, tandis qu'ils s'attendaient à ce qu'un libéral typique donne un score entre 6 ou 7. La différence réelle (1,3 points de différence) est bien moindre que la différence attendue (entre 4 ou 5 points). Même s'il existe des différences certaines entre la droite et la gauche sur la question du mérite,

⁶⁵ Lisa FARWELL et Bernard WEINER, « Bleeding Hearts and the Heartless: Popular Perceptions of Liberal and Conservative Ideologies », *Personality and Social Psychology Bulletin*, 2000, vol. 26, n° 7, p. 845-852.

⁶⁶ Les auteurs se sont également servis d'autres conditions expérimentales, impliquant par exemple des personnes atteintes du SIDA ou obèses, et ont obtenu les mêmes résultats avec ces différentes méthodes.

le consensus est beaucoup plus important que ce que pourrait nous faire croire une trop grande attention portée aux différences.

Le fait qu'il existe un consensus plus fort qu'attendu sur la question du mérite est renforcé par une étude interculturelle réalisée par Lene Aarøe et Michael Bang Petersen⁶⁷. Les auteurs ont comparé les opinions publiques au Danemark et aux États-Unis relatives à l'assurance chômage, en se basant sur un échantillon de 1000 personnes dans chacun des deux pays. Les deux nations se caractérisent par une culture politique extrêmement différente, l'assurance chômage étant très faible aux États-Unis, et beaucoup plus généreuse au Danemark. Plus généralement, le Danemark possède un État-providence très fort, tandis que les États-Unis ont, naturellement, une tradition beaucoup plus libérale (au sens classique du terme). Cependant, les chercheurs ont révélé que les attentes des populations au sein des deux pays vis-à-vis des chômeurs n'étaient pas si différentes.

Les auteurs ont d'abord constaté que les stéréotypes associés aux chômeurs étaient très différents dans les deux pays : les Américains ont beaucoup plus de chance de considérer que les chômeurs sont paresseux, tandis que les Danois ont beaucoup plus de chance de considérer que les chômeurs sont malchanceux. En conséquence, lorsque les chercheurs ont demandé aux populations des deux pays dans quelle mesure ils étaient d'accord avec une politique qui conduirait à un contrôle accru des exigences pour obtenir l'allocation chômage, il a existé une différence assez nette, les Américains se montrant plus favorables à un contrôle accru des chômeurs (0,57 contre 0,49, sur une échelle de 0 à 1, un plus grand chiffre indiquant une plus grande opposition).

Cependant, l'existence d'une telle différence entre pays cache un consensus qui existe si l'on rentre dans des cas précis, qui limitent les différences d'interprétation et donc le poids des stéréotypes. Dans deux autres conditions, les chercheurs ont demandé dans quelle mesure les participants étaient d'accord avec l'idée d'un contrôle accru d'un chômeur présentant des caractéristiques spécifiques. Dans un cas, le chômeur était décrit comme ayant perdu son emploi lors d'un accident de travail. Dans un autre, le chômeur était décrit comme préférant vivre des allocations de chômage. Dans chacune de ces deux conditions, les chercheurs n'ont plus observé de différence significative entre les deux pays : aux États-Unis comme au Danemark, il existait un très fort consensus contre le fait d'augmenter le contrôle du chômeur malchanceux, et en faveur d'un contrôle accru du chômeur « paresseux ». Les différences de croyances quant

⁶⁷ Lene AARØE et Michael Bang PETERSEN, « Crowding Out Culture: Scandinavians and Americans Agree on Social Welfare in the Face of Deservingness Cues », *The Journal of Politics*, 2014, vol. 76, n° 3, p. 684-697.

aux caractéristiques empiriques des chômeurs semblent masquer un consensus au niveau des valeurs, en faveur de l'idée qu'une personne qui ne veut pas travailler ne doit pas être aidée.

Jusqu'à présent, les différentes expériences introduites ici ont surtout concerné des situations locales, c'est-à-dire référant à une distribution au sein d'un groupe restreint d'individus. Cependant, l'importance du mérite se voit aussi dans un contexte où l'on demande aux participants qu'ils prennent une vision globale de la société. Dans une série de deux expériences en 1993, Barbara Mellers et ses collègues ont décrit à leurs participants (des étudiants américains) différentes sociétés⁶⁸. Ces sociétés variaient aussi bien par leur richesse économique moyenne (société relativement pauvre ou relativement riche), leur degré d'inégalités, et leur degré de méritocratie (opérationnalisée sous la forme d'une corrélation faible ou élevée, c'est-à-dire de 0,1 ou 0,9, entre les efforts et les talents des individus et leur niveau de revenu). Ils ont ensuite demandé aux participants de noter chaque société en termes de justice, afin de pouvoir déterminer si c'est l'efficacité économique, l'égalité, ou le mérite, qui avait le plus grand poids. Les chercheurs ont alors découvert que les participants accordaient la plus grande importance au mérite, même si l'égalité et l'efficacité étaient également valorisées. Une telle expérience montre que la valorisation du mérite ne se limite pas à des situations locales, mais que les individus accordent également de l'importance au mérite dans un contexte global.

L'importance accordée au mérite dans un contexte global se retrouve aussi dans les enquêtes effectuées auprès de différents pays européens. Michel Forsé et Maxime Parodi se basent ainsi sur le *Social Value Survey*, réalisé en 1999 auprès de 30000 citoyens de 24 pays de l'Union européenne, pour soutenir qu'il existe un consensus très large autour de l'importance du mérite, de la satisfaction des besoins, et de l'égalité⁶⁹. En moyenne, 90% des Européens disent ainsi qu'il est important ou très important d'assurer la satisfaction des besoins de base, 81% disent qu'il est important ou très important de reconnaître les gens selon leur mérite, et 65% indiquent qu'il est important ou très important de réduire les grandes inégalités. Là encore, il existe une forte reconnaissance de l'importance d'une pluralité de critères de justice, et le mérite émerge comme un principe possédant un très fort soutien à travers les différents pays occidentaux.

⁶⁸ Gregory MITCHELL, Philip E. TETLOCK, Barbara A. MELLERS et Lisa D. ORDÓÑEZ, « Judgments of Social Justice: Compromises Between Equality and Efficiency », *Journal of Personality and Social Psychology*, 1993, vol. 65, n° 4, p. 629-639 ; Lisa D. ORDÓÑEZ et Barbara A. MELLERS, « Trade-offs in fairness and preference judgments », in Barbara A. MELLERS et Jonathan BARON (dir.), *Psychological perspectives on justice, Theory and application*, Cambridge University Press., Barbara Mellers and Jonathan Baron, 1993, p.

⁶⁹ Michel FORSE et Maxime PARODI, « Justice distributive », *Revue de l'OFCE*, 2006, no 98, n° 3, p. 213-244.

A travers ce parcours, nous avons vu que, au moins au sein des pays occidentaux, il existe un très fort consensus sur l'importance du mérite en tant que principe de justice. L'existence d'un tel accord au sein de nos sociétés n'est cependant en rien l'évidence d'une portée universelle du mérite. De fait, plusieurs historiens de la philosophie ont insisté sur le fait que la période actuelle montre une insistance très forte sur l'idée de mérite, qui n'était pas évidente dans la période plus égalitaire des trente glorieuses⁷⁰. Il est bien sûr difficile de comparer la place faite au mérite à travers l'histoire occidentale ; il est cependant possible de remarquer qu'on peut trouver des défenses du mérite chez une grande partie des philosophies classiques, aussi bien chez Platon, Aristote, Thomas d'Aquin, Shaftesbury, Hume, Adam Smith, Kant, et bien d'autres. Le mérite n'est en rien une invention récente. Cela ne veut cependant pas dire qu'il posséderait une portée universelle. Je vais passer en revue un certain nombre d'études qui semblent malgré tout indiquer que le mérite possède un attrait bien au-delà des frontières européennes.

La centralité du mérite dans les conceptions ordinaires : comparaisons internationales

Pour mettre à l'épreuve l'hypothèse d'une universalité des jugements de mérite, je vais procéder en deux étapes. Dans un premier temps, je vais présenter les résultats d'études interculturelles qui semblent montrer une différence des jugements de mérite au sein des différents pays, et qui ont pu être interprétées comme indiquant qu'il n'existe pas d'universel en ce qui concerne les jugements de justice en général. A partir d'une nouvelle analyse de données résultant d'enquête internationales, je montrerai pourquoi de telles différences sont exagérées, et pourquoi l'interprétation la plus cohérente est qu'il existe une très forte similarité dans les jugements de mérite au sein des différentes cultures. Dans un deuxième temps, je présenterai des résultats de psychologie développementale qui semblent indiquer un développement précoce d'un sens du mérite dans diverses sociétés.

Les données interculturelles relatives à la perception de la justice sont difficiles à analyser, pour trois raisons. Tout d'abord, le nombre d'études ayant été réalisées à l'international est

⁷⁰ P. SAVIDAN, *Repenser l'égalité des chances*, op. cit. ; Pierre ROSANVALLON, *La Société des égaux*, Paris, Points, 2013.

beaucoup plus faible que le nombre d'études effectuées au sein des pays occidentaux⁷¹. Ce manque d'études est particulièrement criant en ce qui concerne les pays africains, pour lesquels il n'existe que très peu de données. Une deuxième difficulté est que, dans le cas d'études ayant été réalisées, le nombre de participants est souvent relativement faible, restreignant encore davantage la possibilité d'une analyse⁷². Enfin, les résultats se contentent souvent de montrer une différence entre les populations de deux pays, sans qu'il soit forcément simple d'interpréter ces différences. Par exemple, en ce qui concerne l'analyse des différences ou des similarités entre pays occidentaux et d'Asie de l'est sur la question du mérite et de l'égalité, les différentes études qui ont été réalisées ont donné des résultats opposés, certaines études laissant penser que les personnes d'Asie de l'est attachent plus d'importance au mérite⁷³, d'autres indiquant que les personnes d'Asie de l'est accordent plus d'importance à l'égalité que les habitants des pays occidentaux⁷⁴.

Certains traits relativement stables émergent cependant. De façon générale, ces études tendent plutôt à indiquer une très forte convergence des idées autour de l'importance des critères de productivité et d'effort dans la détermination d'un juste salaire⁷⁵, en ce que l'ensemble des cultures étudiées (en Inde ou en Asie de l'Est comme dans les pays occidentaux) considèrent qu'il s'agit de critères pertinents. Une autre conclusion souvent retrouvée porte sur la question du besoin, c'est-à-dire de savoir dans quelle mesure le salaire d'un individu doit tenir compte des besoins d'un individu ou de sa famille. Plusieurs études tendent à montrer que des participants Indiens attachent plus d'importance au critère du besoin dans la détermination du salaire que des participants américains⁷⁶.

⁷¹ Ceci correspond à la critique du fait que les études en psychologie se concentrent sur un public « W.E.I.R.D. », « étrange », c'est-à-dire occidental (« western »), éduqué, industrialisé, riche et démocratique. Joseph HENRICH, Steven J. HEINE et Ara NORENZAYAN, « The weirdest people in the world? », *The Behavioral and Brain Sciences*, 2010, vol. 33, n° 2-3, p. 61-83; discussion 83-135.

⁷² Pour plus de détails, voir les tailles d'échantillon reportées dans la méta-analyse de Ronald FISCHER et Peter B. SMITH, « Reward Allocation and Culture: A Meta-Analysis », *Journal of Cross-Cultural Psychology*, 2003, vol. 34, n° 3, p. 251-268.

⁷³ Chao C. CHEN, « New Trends in Rewards Allocation Preferences: A Sino-U.S. Comparison », *Academy of Management Journal*, 1995, vol. 38, n° 2, p. 408-428.

⁷⁴ Irwin MAHLER, Lisa GREENBERG et Haruo HAYASHI, « A comparative study of rules of justice: Japanese versus American », *Psychologia: An International Journal of Psychology in the Orient*, 1981, vol. 24, n° 1, p. 1-8.

⁷⁵ Pour une défense de cette interprétation et une présentation des résultats des études inter-culturelles jusqu'à la fin des années 2000, voir le résumé de David Miller dans *Justice for Earthlings: Essays in Political Philosophy*, Cambridge, Cambridge University Press, 2013.

⁷⁶ Virginia MURPHY-BERMAN, John J. BERMAN, Purnima SINGH, Anju PACHAURI et Pramod KUMAR, « Factors affecting allocation to needy and meritorious recipients: A cross-cultural comparison », *Journal of Personality and Social Psychology*, 1984, vol. 46, n° 6, p. 1267-1272 ; JOHN J. BERMAN, VIRGINIA MURPHY-BERMAN et PURNIMA SINGH, « Cross-Cultural Similarities and Differences in Perceptions of Fairness », *Journal of Cross-Cultural Psychology*, 1985, vol. 16, n° 1, p. 55-67.

En ce qui concerne l'importance de la contribution, des efforts, et des besoins dans la détermination de l'importance du salaire, nous disposons cependant de données d'excellente qualité issues de l'enquête *Social Inequality IV* de l'*International Social Survey Programme*⁷⁷. Pour obtenir une perspective d'ensemble, et faire usage de données de très bonne qualité, j'ai effectué une analyse de questions posées dans le cadre de cette enquête à des populations de 41 pays, incluant un certain nombre de pays non-occidentaux.

Une réanalyse de l'enquête Social Inequality IV

Avant d'entrer dans les détails de l'enquête, un mot sur la méthode employée. L'enquête *Social Inequality IV* a recruté plus de 1000 participants dans chacun des 41 pays concernés. Dans tous les cas, le recrutement de l'échantillon s'est fait selon une méthode probabiliste, indiquant que l'échantillon obtenu est assez largement représentatif de la population concernée. Il s'agit donc de données d'une qualité remarquable. Au niveau géographique, les pays concernés incluent 27 pays européens (d'Europe de l'est et de l'ouest), 5 pays asiatiques (le Japon, la Corée, la Chine, Taiwan, et les Philippines), trois pays d'Amérique du Sud (le Chili, l'Argentine, et le Venezuela), et un ensemble de pays d'autres origines incluant plusieurs pays anglo-saxons (les États-Unis, l'Australie), l'Afrique du Sud, la Turquie et Israël. Les pays occidentaux sont donc dominants dans cet ensemble, mais l'enquête est néanmoins remarquable par son inclusion d'un assez large ensemble de cultures différentes.

Cette enquête contient une question qui nous intéresse directement. Les enquêteurs ont posé à chaque participant la question suivante : « À votre avis, pour décider de ce que quelqu'un doit gagner, quelle importance chacun des éléments suivants devrait avoir ? », suivi d'une liste de 6 éléments différents : « Les responsabilités dans le travail », « Le nombre d'années passées à étudier et se former », « Ce qui est nécessaire pour faire vivre une famille », « La présence d'enfants à charge », « Le degré auquel le travail est bien fait », et enfin « Le fait de travailler dur ».

Pour chaque élément, le participant avait le choix entre cinq types de réponse :

« C'est essentiel »

« C'est très important »

⁷⁷ ISSP Research Group (2017): International Social Survey Programme: Social Inequality IV. ISSP 2009. GESIS Data Archive, Cologne. ZA5400 Data file Version 4.0.0, doi: 10.4232/1.12777.

« C'est plutôt important »

« Ce n'est pas très important »

« Cela n'a aucune importance »

Les différents facteurs proposés à l'attention des participants sont particulièrement intéressants, parce qu'ils combinent un mélange de critères qui sont considérés traditionnellement comme étant liés au mérite (le fait d'avoir des responsabilités, de faire des efforts, de faire bien son travail) et des critères liés aux besoins⁷⁸ (comme ce qui est nécessaire pour faire vivre sa famille, ou le fait d'avoir des enfants). Ces données permettent donc de tester directement dans quelle mesure les critères liés au mérite sont considérés comme pertinents à l'international, et dans quelle mesure il existe effectivement de fortes variations dans la prise en compte des besoins individuels pour déterminer le juste salaire que doit recevoir un individu.

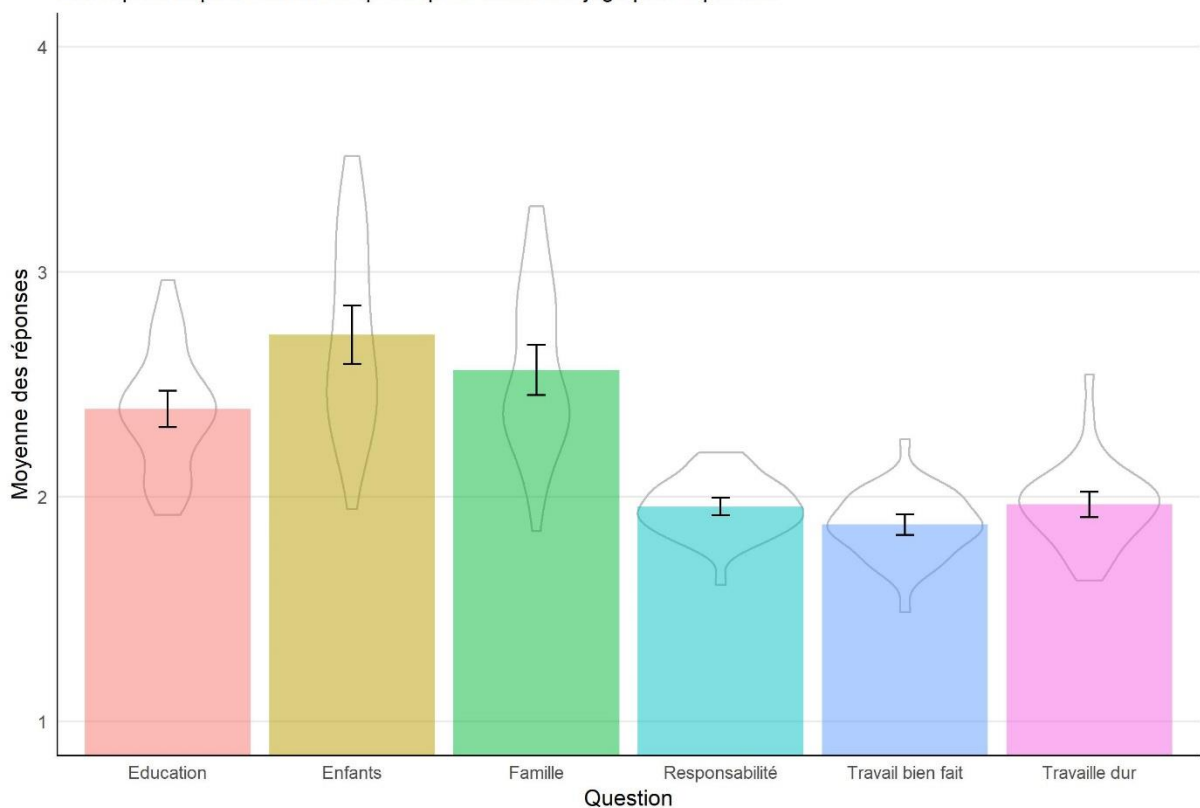
Même si d'autres aspects de l'enquête *Social Inequality IV* ont été très bien exploités dans de nombreuses publications, à ma connaissance, aucun travail n'a jusqu'à présent analysé les réponses à ces questions en particulier. Ce manque d'exploitation est d'autant plus surprenant que l'analyse de ces données révèlent plusieurs résultats forts.

Le premier résultat, le plus évident, est que les six facteurs mentionnés semblent globalement pertinents pour les populations concernées. Reprenant les codes utilisés dans le cadre de l'enquête, j'ai attribué la valeur 1 à la réponse « C'est essentiel », la valeur 2 à « C'est important », et ainsi de suite jusqu'à attribuer la valeur 5 à la réponse « Cela n'a aucune importance », de telle sorte que les réponses plus basses indiquent un plus fort attachement au critère. Si l'on prend la moyenne des réponses au sein de chaque pays, et que l'on calcule ensuite la moyenne des résultats obtenus entre les différents pays, alors les six facteurs sont bien considérés comme étant « très importants » ou « plutôt importants » (voir la Figure 1). Même le critère considéré comme globalement le moins pertinent (« La présence d'enfants à charge ») est tout de même considéré comme étant plutôt important au sein des différents pays, avec une moyenne globale de 2,72. Les personnes interrogées au sein des différents pays semblent donc prendre en compte aussi bien des facteurs qui sont directement liés au travail et au mérite,

⁷⁸ Le critère lié au fait d'avoir fait de longues études est un peu particulier, en ce qu'il n'offre pas d'interprétation facile. Comme l'ont bien montré plusieurs enquêtes, la longueur des études est un critère ambigu de mérite, en tant qu'il n'est pas toujours évident qu'une personne ayant fait des études est toujours capable d'effectuer un travail de meilleure qualité qu'une personne ayant fait des études moins longues. Il semble ainsi exister un certain consensus sur le fait qu'une personne ayant fait des études mérite de gagner davantage si ces études l'aident à être plus productive, mais pas si cette personne n'est pas plus efficace au travail que si elle n'avait fait aucune étude. Cf. par exemple l'enquête d'Elise Tenret sur le rapport des étudiants au mérite. Elise TENRET, *Les étudiants et le mérite - À quoi bon être diplômé ?*, Paris, La Documentation Française, 2011. Pour contourner ces difficultés, je me concentre par la suite sur les autres critères, qui sont plus facilement interprétables en termes de mérite ou de besoins.

comme le fait de bien faire son travail, que des facteurs qui semblent à première vue lui être extérieurs et être davantage liés à la question du besoin, comme le fait de savoir si la personne a des enfants à charge.

Fig. 1. Moyenne des réponses entre les différents pays.
Des réponses plus basses indiquent qu'un critère est jugé plus important.



Le deuxième résultat crucial, cependant, est que tous les facteurs ne sont pas considérés comme étant aussi importants. D'une manière assez claire, les facteurs considérés comme étant les plus importants sont ceux qui sont directement liés au travail, comme le fait de savoir si le travail implique des responsabilités (moyenne de 1,96), le fait de savoir si la personne fait bien son travail (moyenne de 1,88), et le fait de savoir si la personne fait beaucoup d'efforts (moyenne de 1,97). Les critères les moins pertinents sont ceux liés aux besoins (moyenne de 2,56 dans le cas de la famille, et de 2,72 dans le cas de savoir si la personne a des enfants). Ces différences sont hautement significatives⁷⁹. Les critères liés au mérite semblent donc bien les plus pertinents pour l'ensemble des pays.

⁷⁹ J'ai effectué une série de t-tests appariés pour comparer les réponses moyennes. Pour simplifier, je me concentre sur le cas de la question portant sur ce qui est nécessaire pour faire vivre une famille, puisque c'est cette réponse qui a obtenu la plus forte adhésion par rapport à la question liée au fait d'avoir des enfants. Même dans le cas des questions liées à la famille, les réponses sont significativement plus hautes (indiquant un critère jugé moins important) que pour toutes les questions liées au mérite. La différence avec la question portant sur le fait de bien travailler est : $M_d = -0,68$, 95% CI $[-0,80, -0,56]$, $t(39) = -11,26$, $p < ,001$. La différence avec le fait de travailler dur est : $M_d = -0,60$, 95% CI $[-0,71, -0,49]$, $t(40) = -10,78$, $p < ,001$. Enfin, la différence avec le fait d'avoir des responsabilités est : $M_d = -0,61$, 95% CI $[-0,73, -0,49]$, $t(40) = -10,25$, $p < ,001$.

Le résultat le plus intéressant tient aux différences de consensus des critères liés au mérite par rapport aux critères liés aux besoins. En un mot, les critères liés au mérite font consensus aussi bien au sein de chaque pays qu'entre les différents pays. Il existe très peu de variations entre les différents pays quant à l'appréciation des trois critères liés aux responsabilités, au travail bien fait, et au fait de faire beaucoup d'efforts. A l'inverse, les critères liés au besoin font l'objet de beaucoup plus de divisions et de variations, aussi bien entre pays qu'à l'intérieur de chaque pays.

Dans le cas des critères liés au mérite, pour donner une idée du manque de variabilité entre pays, j'ai représenté dans la figure 2 un graphe de Cleveland représentant les scores des différents pays à la question relative à l'importance du travail bien fait. Le lecteur peut voir que, en dehors de la Bulgarie et du Danemark, toutes les réponses se trouvent entre 1,65 et 2,1. Même dans les cas les plus extrêmes, les réponses sont au moins autour de 1,5 (dans le cas de la Bulgarie) et 2,25 (dans le cas du Danemark). L'écart-type des moyennes entre les pays est de 0,143, ce qui est extrêmement faible. Il existe bien une variation entre pays, certes, mais elle reste limitée en comparaison avec l'uniformité qui semble régner sur ces questions. Le lecteur intéressé pourra trouver dans l'appendice à ce chapitre les graphes de Cleveland pour les deux autres questions liées au mérite, c'est-à-dire pour les questions sur l'importance de la responsabilité et des efforts.

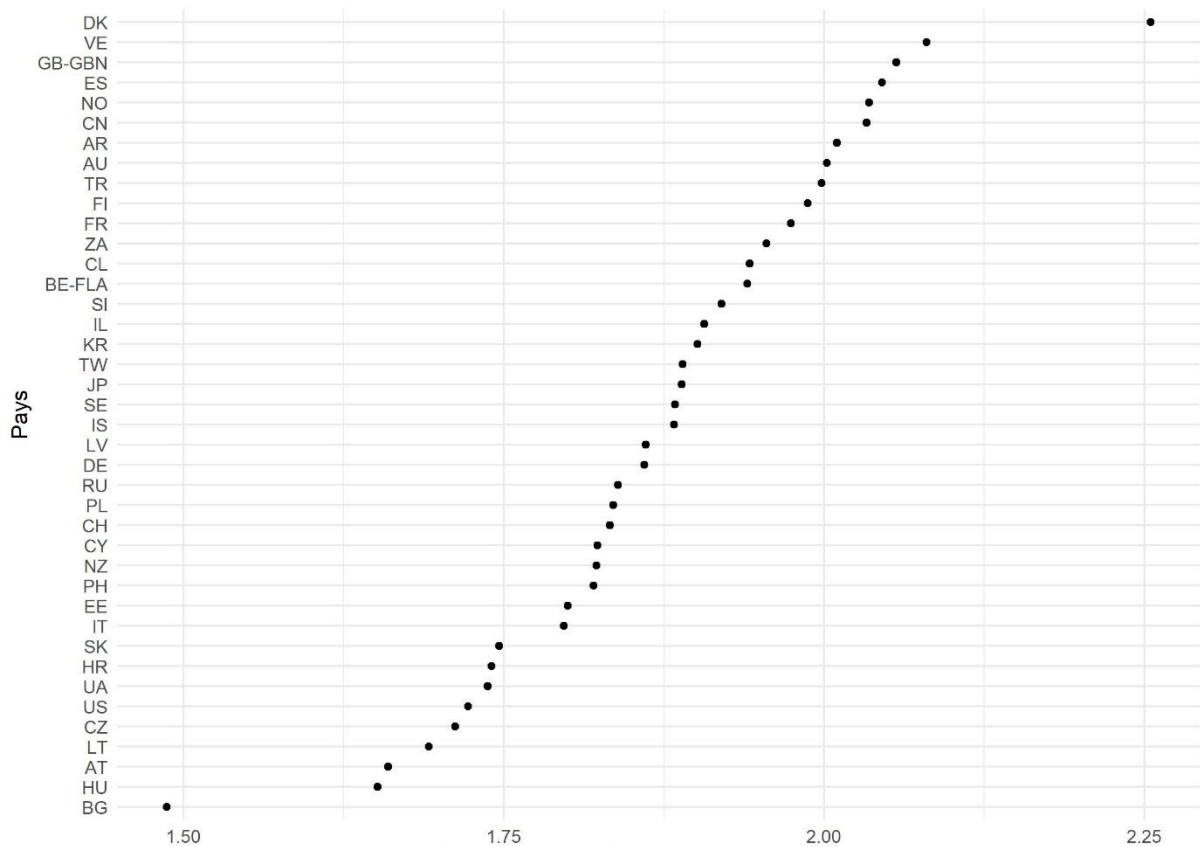


Fig. 2. Moyenne des réponses à la question liée à l'importance du travail bien fait pour déterminer le salaire.

Cette remarquable homogénéité interculturelle est redoublée par une très faible variation également visible à l'intérieur de chaque pays. Dans l'intégralité des pays, la très forte majorité des participants choisit d'indiquer les facteurs liés au mérite comme étant « essentiel », « très important » ou « plutôt important ». Par exemple, dans le cas de l'importance du travail bien fait, dans tous les pays, au moins 93% de la population a fait l'un de ces trois choix. Dans 35 des 40 pays⁸⁰, la réponse modale était que le travail bien fait était un critère « très important ». Pour les 5 pays restants, le travail bien fait était décrit comme un critère « essentiel ». A chaque fois, moins d'un pour cent des personnes interrogées avait désigné le travail bien fait comme un critère n'ayant « aucune importance ». L'analyse des données portant sur les critères des efforts et de la responsabilité donnent des résultats similaires. Les résultats sont donc sans appel : dans tous les pays, pour l'immense majorité de la population, les critères liés au mérite apparaissent bien comme des critères décisifs pour déterminer le salaire d'un individu. Il

⁸⁰ Étonnamment, cette question n'a pas été posée au Portugal, ce qui fait que l'on ne possède que 40 pays pour étudier cette question en particulier.

n'existe donc qu'une très faible variation sur les jugements de l'importance de ces critères, aussi bien entre pays qu'à l'intérieur de chaque pays.

Passons maintenant à l'analyse des critères liés aux besoins. Pour les deux critères « Ce qui est nécessaire pour faire vivre une famille » et « La présence d'enfants à charge », la variation culturelle se révèle être beaucoup plus importante, et il existe moins de consensus au sein des différents pays. On peut prouver cela de façon simple en comparant les différences de variances (à l'international et au sein de chaque pays) pour les questions liées au mérite, d'un côté, et pour les questions liées aux besoins, de l'autre. L'écart-type indiquant la variation des moyennes entre les différents pays est de 0,13 dans le cas du critère de responsabilité, 0,14 dans le cas du travail bien fait, 0,18 dans le cas des efforts. Dans le cas des critères liés aux besoins, l'écart-type est deux fois plus important, passant à 0,36 dans le cas des ressources nécessaires pour subvenir à sa famille, et même à 0,41 dans le cas du fait d'avoir des enfants. Ces différences sont hautement significatives⁸¹.

Au sein des différents pays, l'écart-type est également plus important pour les questions liées aux besoins que pour les questions liées au mérite, même si la différence entre les deux types de critères est moins importante que dans le cas des comparaisons entre pays. L'écart-type à l'intérieur de chaque pays tourne autour de 1 dans le cas des questions liées aux besoins (1,03 dans le cas du fait d'avoir des enfants, 0,97 pour le critère lié à la famille), alors que l'écart-type est en moyenne autour de 0,75 pour les questions liées au mérite. Dans presque tous les pays, l'écart-type est significativement plus important dans le cas des questions liées aux besoins que dans le cas des questions liées au mérite⁸². Les critères liés aux besoins sont donc davantage controversés à l'intérieur de chaque pays, et montrent donc bien plus de variations à l'international.

Pour donner une idée plus précise des variations liées aux critères de la famille et des enfants, la figure 3 représente un graphe de Cleveland montrant les différentes moyennes des pays pour

⁸¹ En utilisant un F-test pour comparer les variances dans le cas des efforts et le cas du fait de pouvoir subvenir à sa famille, on obtient par exemple $F(40, 40) = 3,95$, $p < .0001$, indiquant une plus faible variance dans les réponses liées à l'importance des efforts. Or, il s'agit de la différence la plus faible entre un critère lié au mérite et un critère lié aux besoins ; toutes les autres différences sont encore plus importantes, conduisant à des p-values encore plus faibles.

⁸² Par exemple, on peut comparer la différence de variance entre les réponses données à la question liée à l'importance de la famille, et à l'importance du fait de travailler dur. Dans 40 des 41 pays considérés, la variance des réponses liées à l'importance du fait de travailler dur est significativement plus faible que la variance des réponses liées à la famille, en se servant d'un F-test pour comparer les variances. Le seul pays où la variance est légèrement plus élevée dans le cas des réponses liées à l'importance de travailler dur est le Venezuela. Ces résultats doivent cependant être interprétés avec prudence : les réponses ne sont pas normalement distribuées, et l'usage d'un F-test n'est pas approprié dans ce cas. Il n'en reste pas moins remarquable que la variance soit considérablement plus faible pour les questions liées au mérite dans 40 des 41 pays.

le critère « Ce qui est nécessaire pour faire vivre une famille ». Comme le lecteur peut le remarquer, les moyennes vont de 1,8 dans le cas des Philippines à 3,3 dans le cas de l'Estonie, ce qui constitue un écart maximal de moyenne bien plus important que dans le cas des critères liés au mérite. Cependant, malgré ce fort écart, il faut bien sûr remarquer que les moyennes sont loin de couvrir tout l'éventail des jugements possibles. De manière générale, la plupart des moyennes se situent entre 2,2 et 2,9, indiquant une assez forte convergence autour de l'idée que ces critères liés aux besoins sont importants, mais moins que les critères liés au mérite. Le graphe de Cleveland portant sur le critère lié au fait d'avoir des enfants, reproduit dans l'appendice, révèle une variation similaire des moyennes.

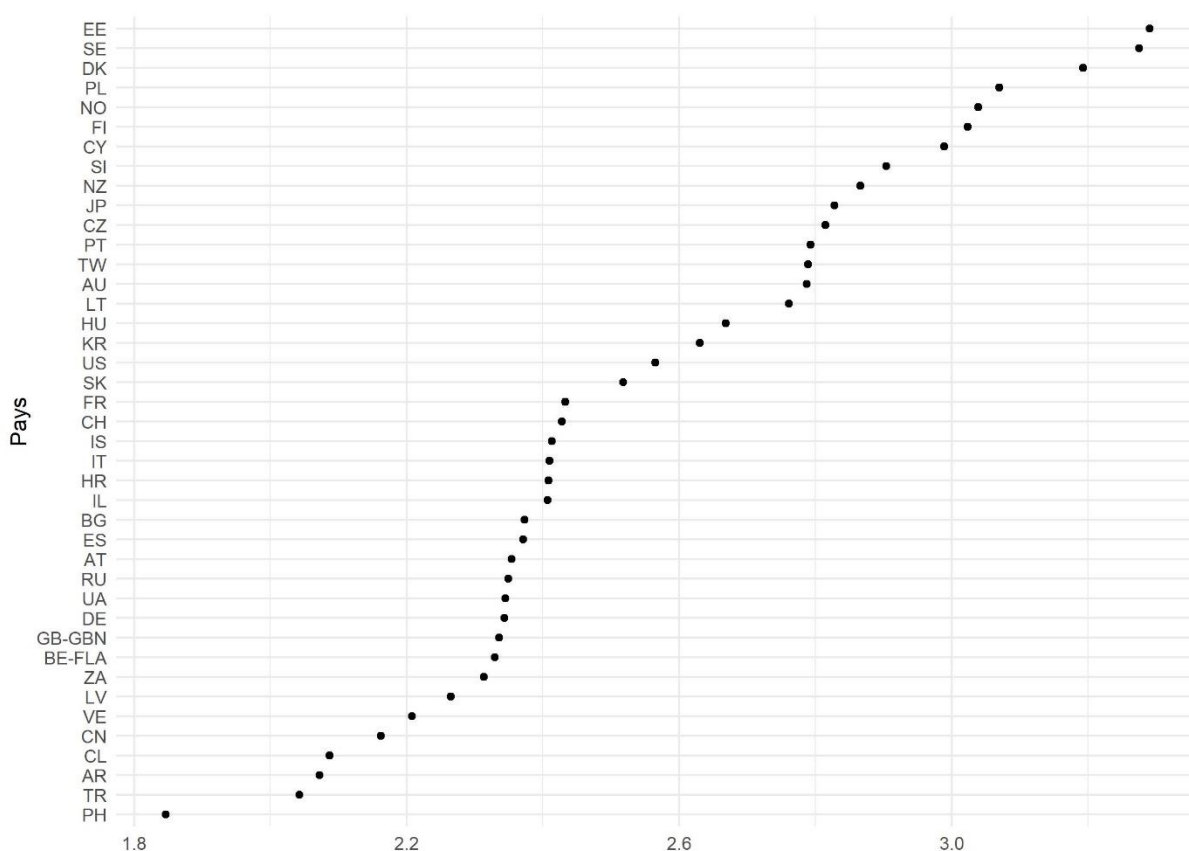


Fig. 3. Réponses à la question sur l'importance du fait de subvenir aux besoins de sa famille.

L'analyse des données au sein de chaque pays révèle également qu'il existe une plus grande variation que dans le cas des critères liés au mérite. En moyenne, 18% des enquêtés dans chaque pays indiquaient « Ce n'est pas très important » ou « Cela n'a aucune importance » en ce qui concerne le critère de ce qui est nécessaire pour subvenir aux besoins de la famille. Presque aucune personne interrogée n'avait choisi ces réponses dans le cas des critères liés au mérite. Dans le cas de l'Estonie, le pays avec le plus faible taux de participants considérant le critère

des besoins de la famille comme important, 44% de la population a indiqué que ce n'était pas très important, ou que cela n'avait aucune importance. Même si, dans la plupart des pays, il existe une forte majorité considérant que ces critères sont au moins plutôt importants, il existe une minorité relativement forte qui ne considère pas ces critères comme pertinents.

Que faut-il conclure de ces données sur l'universalité des conceptions du mérite et de la justice sociale ? Tout d'abord, il semble que l'importance du travail bien fait, de la prise de responsabilité, et des efforts relèvent d'un très large consensus dans un très grand nombre de cultures différentes. Il est difficile de conclure à l'universalité de ces pratiques, cependant, étant donné que très peu de pays africains sont inclus, et qu'aucune société de petite taille (qu'il s'agisse de chasseurs-cueilleurs ou d'horticulteurs) n'a été interrogée. Toutefois, il s'agit de données de bonne qualité indiquant que le soutien pour les idées apparentées au mérite n'est pas réduit aux populations des pays occidentaux.

Ensuite, il semble que l'importance de la satisfaction des besoins dans la détermination du salaire reflète une beaucoup plus grande diversité d'intuitions, que ce soit entre cultures ou au sein de chaque société. Une telle diversité pourrait refléter l'existence de désaccords culturels forts autour de l'importance d'un principe de justice qui pourrait sembler fondamental, à savoir l'idée qu'il est bon de satisfaire les besoins des travailleurs au sein d'une société.

Conclure à l'existence d'une différence fondamentale dans la conception de la justice serait tout à fait prématuré, cependant. En effet, ces différences dans la prise en compte du besoin sur le marché du travail peuvent relever de plusieurs explications différentes. L'explication compatible avec la plus forte variation culturelle pourrait être le fait que certaines cultures attachent plus de valeur au fait que l'ensemble des citoyens aient leurs besoins satisfaits. C'est notamment une explication de ce type qui a souvent été invoquée par les recherches en psychologie culturelle qui ont montré que les Indiens, par exemple, considéraient davantage que les Américains que les besoins des individus étaient un critère pertinent pour déterminer le salaire d'un individu⁸³. Une tout autre explication est cependant possible, compatible avec une forte universalité d'un attachement aussi bien au mérite qu'à la satisfaction des besoins.

En effet, si l'on regarde les pays qui considèrent la prise en compte des besoins comme le moins légitime sur le marché du travail, on trouve un certain nombre de pays, comme la Suède ou le Danemark, qui disposent d'un État-Providence extrêmement développé, et qui ont agi fortement pour développer un système capable de satisfaire les besoins de l'ensemble des

⁸³ V. MURPHY-BERMAN, J. J. BERMAN, P. SINGH, A. PACHAURI et P. KUMAR, « Factors affecting allocation to needy and meritorious recipients », *op. cit.* ; JOHN J. BERMAN, VIRGINIA MURPHY-BERMAN et PURNIMA SINGH, « Cross-Cultural Similarities and Differences in Perceptions of Fairness », *op. cit.*

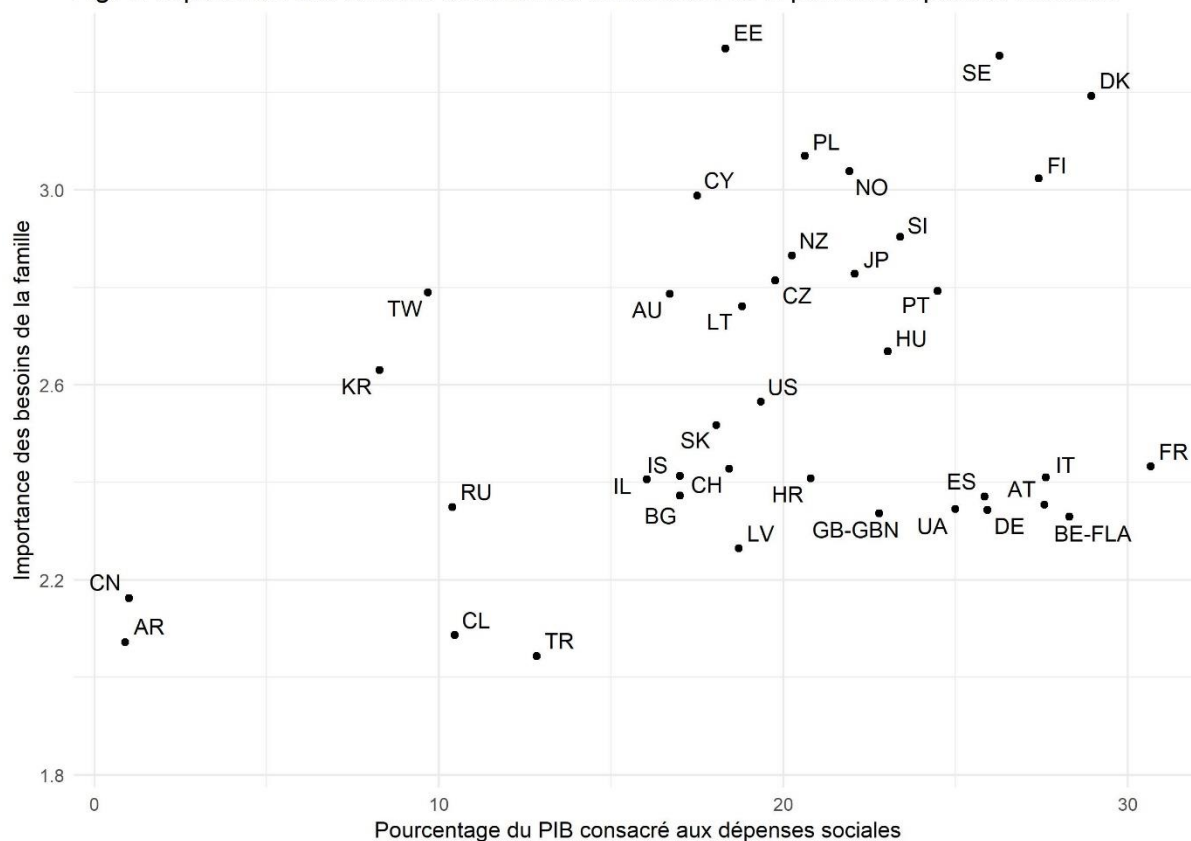
citoyens. Dans ce cadre, il est possible de formuler une autre hypothèse : le fait de ne pas considérer la prise en compte des besoins comme pertinents sur le marché du travail ne refléterait pas un attachement culturel moindre à la satisfaction des besoins des citoyens, mais refléterait au contraire le fait qu'il existe d'autres systèmes de protection au sein de la société qui prennent déjà en charge la satisfaction des besoins. En d'autres termes, il serait d'autant moins important pour les entreprises de prendre en compte le besoin de leurs salariés que ces besoins auraient déjà été pris en compte par l'État, ou par d'autres formes d'organisation sociale.

Pour prendre un exemple concret, on peut en effet considérer qu'il est moins pertinent de prendre en compte le fait qu'une personne a des enfants pour déterminer son salaire si cette personne reçoit déjà des allocations familiales, qui amortissent les coûts associés au fait d'avoir des enfants. Les variations entre pays dans les réponses à l'enquête *Social Inequality IV* ne refléteraient donc pas des différences culturelles de valeur, mais simplement des différences dans les systèmes de protection sociale à l'œuvre dans les différents pays.

Pour tester de façon plus rigoureuse cette idée, j'ai effectué une analyse corrélacionnelle comparant la part des dépenses sociales publiques au sein de chaque pays, avec les réponses relatives à l'importance des besoins dans la famille ou du fait d'avoir des enfants. Les données relatives à la part des dépenses sociales publiques sont tirées des données de l'OCDE et de l'organisation internationale du travail⁸⁴. La figure 4 représente un nuage de points comparant la part des dépenses sociales dans les différents pays avec les moyennes des réponses à la question relative à l'importance de satisfaire les besoins de la famille. Il existe une corrélation positive et significative entre les deux valeurs ($r = 0,35$, $t(36) = 2,23$, $p = ,03$). Une comparaison entre les dépenses sociales et les réponses portant sur le critère du fait d'avoir des enfants donne des résultats similaires ($r = 0,29$, $t(36) = 1,81$, $p = ,08$). Ces corrélations sont compatibles avec l'idée que les populations considérant que les besoins ne sont pas pertinents sur le marché du travail le font parce qu'ils considèrent que les besoins sont déjà pris en charge par les structures étatiques. Dans ce cadre, on pourrait considérer que les divergences de réponses aux questions de *Social Inequality IV* masquent une universalité des valeurs, les participants de tous les pays considérant qu'une société juste est une société capable à la fois de satisfaire les besoins et de récompenser le mérite.

⁸⁴ Pour la plupart des pays, les données sur les dépenses sociales sont pour l'année 2010 ; dans certains cas, l'année 2010 n'était pas disponible, et j'ai sélectionné une année précédente. Ni l'OCDE ni l'OIT ne disposent de données pour le Venezuela, pour l'Afrique du Sud, et pour les Philippines, donc ces trois pays sont exclus de l'analyse.

Fig. 4. Importance des besoins de la famille en fonction de la part des dépenses sociales.



Il faut naturellement rester prudent dans l'analyse de ces corrélations, et ce, pour plusieurs raisons. Tout d'abord, l'échantillon est très petit (une quarantaine de pays), ce qui implique que le niveau réel de corrélation entre l'étendue des dépenses sociales et les perceptions du besoin sur le marché du travail pourrait tout à fait être très faible, ou très important ; l'intervalle de confiance des corrélations inclut, dans notre cas, des corrélations très faibles et très fortes (entre 0 et 0,6). Par ailleurs, les différents pays de notre échantillon ne sont pas indépendants : les pays du Nord de l'Europe, comme la Suède, la Finlande et le Danemark comptent comme trois pays différents dans notre corrélation, mais possèdent en réalité de fortes similarités dans leur mode de vie et dans leur système social. Pour toutes ces raisons, ces données doivent seulement être considérées comme suggestives, en tant qu'elles vont dans la direction d'une universalité de la perception de l'importance des besoins et du mérite, sans pour autant être décisives.

Cette analyse de données internationales nous a donc permis de présenter des données suggérant une certaine universalité des jugements de mérite. Même si la prise en compte des critères de besoins sur le marché du travail a montré beaucoup plus de variabilités entre pays,

ces données restent compatibles avec une simple différence de systèmes de protection sociale, et non pas une différence nette de valeurs entre pays. Avant de conclure ce chapitre, je vais présenter un dernier facteur indiquant une forte similarité des jugements de mérite entre différents pays : les études en psychologie du développement étudiant les jugements moraux de jeunes enfants dans différents pays.

Une similarité dans les conceptions du mérite qui se retrouve chez de jeunes enfants

Le fait que le mérite est extrêmement populaire dans de nombreuses sociétés demande une explication, et plusieurs possibilités sont envisageables. Il ne s'agira pas ici de trancher cette question, ce qui demanderait évidemment une somme d'enquêtes et d'études qui ne sont pas disponibles à l'heure actuelle, et dépassent le cadre de cette thèse. Cependant, il reste important d'envisager plusieurs hypothèses, et de voir dans quelle mesure des données empiriques permettent d'éliminer des possibilités peu probables. Plus précisément, il faut envisager la question de savoir si le mérite fait partie de la nature de l'homme, ou serait simplement le fruit de pratiques culturelles très largement répandues⁸⁵.

On peut tout d'abord penser à la possibilité que des pratiques de mérite soient très largement partagées à cause du fait que le mérite est ancré dans la nature humaine, à la suite de deux processus évolutionnistes. Si l'on adopte une perspective d'évolution biologique, on peut penser au fait que l'idée de mérite soit une adaptation directe à l'environnement ancestral. Ceci voudrait dire que la faculté à former des jugements de mérite aurait conféré un avantage en termes de reproduction de survie dans l'environnement ancestral, par exemple en facilitant la coopération au sein du groupe.

Si l'on ne pense pas que le mérite est une adaptation directe, on peut quand même formuler une deuxième hypothèse évolutionniste. On peut penser à la possibilité que les jugements de mérite sont une conséquence secondaire d'adaptations biologiques, même s'ils ne seraient pas une adaptation directe. Par exemple, étant donné que les pratiques de réciprocité ont une fonction biologique claire, et se retrouvent chez de nombreux animaux, y compris chez les primates, on peut faire l'hypothèse que le mérite est simplement la généralisation et la

⁸⁵ Evidemment, même si l'on ne considère pas que le mérite fasse partie de la nature de l'homme, il resterait à expliquer pourquoi il se retrouve dans un aussi grand nombre de cultures.

rationalisation de ces pratiques dans une visée universelle. En un mot, on pourrait faire l'hypothèse que les jugements de mérite seraient la conséquence du fait que les êtres humains valorisent la réciprocité, et éprouvent une sorte de gratitude ou de bienfaisance quand ils voient quelqu'un effectuer une action positive. Le mérite serait alors un type de « gratitude généralisée », comme en ont fait l'hypothèse Adam Smith ou Sidgwick.

Je mentionne ces deux possibilités d'une évolution naturelle des jugements de mérite pour la simple raison que ces théories peuvent se soumettre à un test empirique, basé sur les travaux en psychologie développementale. En effet, l'hypothèse d'une explication fondée sur la nature humaine gagne en plausibilité si l'on peut trouver des traces des jugements de mérite même à un jeune âge. C'est pour cette raison que plusieurs psychologues évolutionnistes, notamment Coralie Chevalier et Nicolas Baumard, ont cherché récemment à étudier la compréhension des enfants du mérite dans plusieurs sociétés, à la fois occidentales et non-occidentales. Les recherches sont pour l'instant limitées, et ne permettent certainement pas de trancher dans quelle mesure le mérite doit être compris comme un trait culturel ou naturel. Cependant, elles montrent déjà qu'une compréhension du mérite intervient assez tôt dans un certain nombre de cultures différentes.

L'équipe conduite par Coralie Chevalier et Nicolas Baumard a testé le même type d'expériences auprès d'enfants âgés de 3 à 5 ans dans quatre cultures différentes : des enfants dans des écoles en France, en Chine, au Japon, et chez les Turkana du Kenya⁸⁶. Comme c'est souvent le cas à la fois en psychologie culturelle et en psychologie du développement, les échantillons étaient faibles : 75 enfants ont été testés en France, mais seulement 39 au Japon, 42 en Chine, et 20 au Kenya. Cependant, malgré ce faible nombre de participants, les résultats sont assez convaincants, la taille d'effet ayant été très importante au sein de chaque groupe.

Avant de rentrer dans le détail des expériences, il faut dire un mot sur des résultats solides obtenus en psychologie du développement moral. De nombreuses expériences ont montré que les enfants dans les pays occidentaux préfèrent, à partir de 3 ans et ce jusqu'à 9 ou 10 ans, des distributions égalitaires, et non des distributions basées sur le mérite⁸⁷. Il s'agit d'un résultat

⁸⁶ Nicolas BAUMARD, Olivier MASCARO et Coralie CHEVALLIER, « Preschoolers are able to take merit into account when distributing goods », *Developmental Psychology*, 2012, vol. 48, n° 2, p. 492-498 ; Coralie CHEVALLIER, Jing XU, Kuniko ADACHI, Jean-Baptiste van der HENST et Nicolas BAUMARD, « Preschoolers' Understanding of Merit in Two Asian Societies », *PLOS ONE*, 2015, vol. 10, n° 5, p. e0114717 ; P. LIÉNARD, C. CHEVALLIER, O. MASCARO, P. KIURA et N. BAUMARD, « Early Understanding of Merit in Turkana Children », *Journal of Cognition and Culture*, 2013, vol. 13, n° 1-2, p. 57-66.

⁸⁷ Je parle ici de la psychologie du jugement moral, et non du comportement effectif. En général, les jeunes enfants montrent une tendance à « l'égoïsme » (entendu ici comme le simple refus de partage avec d'autres enfants, sans connotation morale) sans commune mesure avec le comportement des adultes, qui ont une tendance beaucoup plus nette à partager. Ce comportement refusant le partage égalitaire est distinct d'un jugement moral qui est très tôt égalitariste.

robuste, qui semble aller à l'encontre de l'idée que le mérite puisse avoir une base naturelle⁸⁸. Coralie Chevalier et ses collègues ont tenté de montrer que, si ce type de résultats montrait en général une préférence pour l'égalité dès un jeune âge, il ne montrait pas que les enfants étaient incapables de prendre le mérite en compte. Autrement dit, les auteurs ont cherché à dépasser la question de la préférence, pour étudier la capacité des enfants à prendre le mérite en compte. Si des explications auxiliaires seraient nécessaires pour expliquer pourquoi une préférence pour le mérite ne se manifeste que vers 9 ou 10 ans, il s'agit d'une première étape pour confirmer la possibilité que le mérite possède une base évolutionnaire adaptative.

Passons maintenant à la description des expériences utilisées par Chevalier et ses collègues. Dans toutes les sociétés étudiées, les vignettes employées décrivaient l'histoire de deux enfants qui ont commencé à préparer des gâteaux ensemble. Après avoir préparé des gâteaux pendant un petit moment, l'un des deux enfants choisit d'aller jouer plutôt que de continuer à cuisiner. Un des deux enfants finit donc de préparer les gâteaux tout seul. A la fin de la préparation, trois gâteaux ont été faits, et c'est alors que l'expérimentateur demande à l'enfant comment il choisirait de distribuer les gâteaux entre les deux enfants. L'expérimentateur précise également qu'il n'est pas possible de diviser un gâteau en deux. Dans un premier temps, l'enfant est libre de distribuer les gâteaux comme il veut, ce qui inclut la possibilité d'effectuer une distribution égalitaire d'un gâteau par personne. Si l'enfant choisit effectivement de ne pas distribuer tous les gâteaux, l'expérimentateur lui demande s'il veut distribuer le gâteau restant à l'un des deux enfants. Cette méthode expérimentale a bien sûr pour but de voir si, quand il est forcé de faire un choix, l'enfant aura bien une préférence pour l'enfant qui a fait le plus d'efforts dans la préparation des gâteaux, comme le prédirait une théorie qui verrait une apparition précoce d'un sens du mérite.

Malgré les variations entre échantillons, les résultats sont globalement concordants entre les différentes cultures. Dans la première phase, quand l'enfant a la possibilité de choisir entre distribuer deux ou trois gâteaux, la plupart des enfants choisissent une distribution égalitaire en attribuant un gâteau à chaque enfant. Dans chaque échantillon, plusieurs enfants choisissent cependant d'attribuer deux des trois gâteaux à l'enfant qui a le plus travaillé, tandis que le choix d'attribuer deux gâteaux à l'enfant qui a choisi de jouer est très rare. Enfin, dans la dernière phase, quand l'expérimentateur offre à l'enfant la possibilité d'attribuer le dernier gâteau, la grande majorité des enfants choisit de le donner à l'enfant qui a le plus travaillé. Si l'on combine les résultats de la première et de la deuxième phase, dans toutes les cultures, une grande majorité

⁸⁸ Je simplifie : de nombreux phénomènes biologiques sont naturels, et dépendent de facteurs génétiques, même s'ils n'interviennent pas avant un certain âge (la puberté est un exemple évident).

des enfants choisit de donner les deux gâteaux à l'enfant ayant le plus travaillé. 56 des 75 enfants ont fait ce choix dans l'échantillon français ; 28 des 39 enfants ont fait de même dans l'échantillon japonais ; 35 des 42 enfants l'ont fait dans l'échantillon chinois ; et 16 des 20 enfants dans l'échantillon du Turkana. Tous ces résultats sont hautement significatifs, ce qui laisse présager des résultats robustes malgré les échantillons de petite taille.

Il faut bien insister sur le fait que l'interprétation de ces résultats est délicate, et qu'ils ne constituent bien évidemment pas une preuve que les jugements de mérite sont le résultat direct d'une adaptation biologique, ni même qu'ils sont naturels ou innés. Il s'agit d'une situation où l'enfant est contraint à faire un choix forcé. De telles expériences montrent seulement que, quand ils sont contraints à faire un choix, des enfants d'un très jeune âge peuvent prendre en compte des critères de mérite comme une raison de donner un bien à quelqu'un. Les fondements de pratiques de récompenses sur la base du mérite semblent donc émerger tôt chez l'enfant, ce qui peut expliquer en partie pourquoi de telles attitudes sont très populaires dans un grand nombre de sociétés. Mais là encore, les données sont limitées : il s'agit seulement de quatre sociétés, avec à chaque fois un échantillon faible. Un plus vaste échantillon, et une diversification du type d'expériences utilisées, seraient nécessaires pour mieux comprendre les contours du développement du mérite.

Conclusion : une idée incontournable dans un très grand nombre de pays

Le mérite possède souvent une image controversée, et a fait l'objet de critiques très vives, notamment au sein du monde universitaire. L'idée de mérite ou méritocratie a été vivement rejetée par un certain nombre de chercheurs en sociologie ou en philosophie. Pour autant, comme nous l'avons vu, l'idée fait l'objet d'un consensus remarquable. Comme nous l'avons observé au sein de la première section de ce chapitre, l'idée de récompenser davantage les gens qui travaillent bien ou font des efforts est une idée extrêmement populaire. Il existe bien des différences entre personnes de différents bords politiques, mais ces différences sont bien moindres que ne le laisseraient penser les stéréotypes, et ces différences tendent à diminuer ou à disparaître dans l'examen de cas précis. De façon générale, il existe un très fort consensus sur ces idées, aussi bien à l'intérieur de chaque société qu'entre sociétés différentes. Enfin, même

si la recherche manque sur ce point, les idées fondamentales du mérite semblent émerger à un âge relativement tôt, dès 3 ou 4 ans, dans un certain nombre de cultures.

Il faut bien évidemment noter une certaine différence dans les jugements ordinaires portant sur ce qu'est un salaire juste. Comme nous l'avons vu, certaines personnes ou certaines cultures semblent porter davantage l'accent sur l'égalité ou sur la satisfaction des besoins, tandis que d'autres insistent davantage sur des critères purement méritocratiques. Ces divisions ne sauraient masquer le fait qu'elles interviennent sur le fond d'un très grand consensus autour de l'idée qu'un travail bien fait, ou qui a fait l'objet de gros efforts, doit être récompensé davantage. Le cœur de l'idée de mérite fait consensus, à droite comme à gauche, en occident comme dans d'autres régions du monde. C'est sur la base de ce consensus que se fera l'analyse de la portée normative du mérite dans les prochains chapitres. Étant donné le caractère central du mérite pour la pensée morale ordinaire, il semble périlleux de supposer qu'une théorie morale plausible pourrait s'en débarrasser. L'importance interculturelle du mérite conduit donc à conférer une certaine plausibilité à un certain conservatisme philosophique, visant à préserver une notion centrale dans nos pratiques quotidiennes.

Appendice au chapitre 3

Graphiques additionnels relatifs à l'analyse des données de l'enquête Social Inequality IV

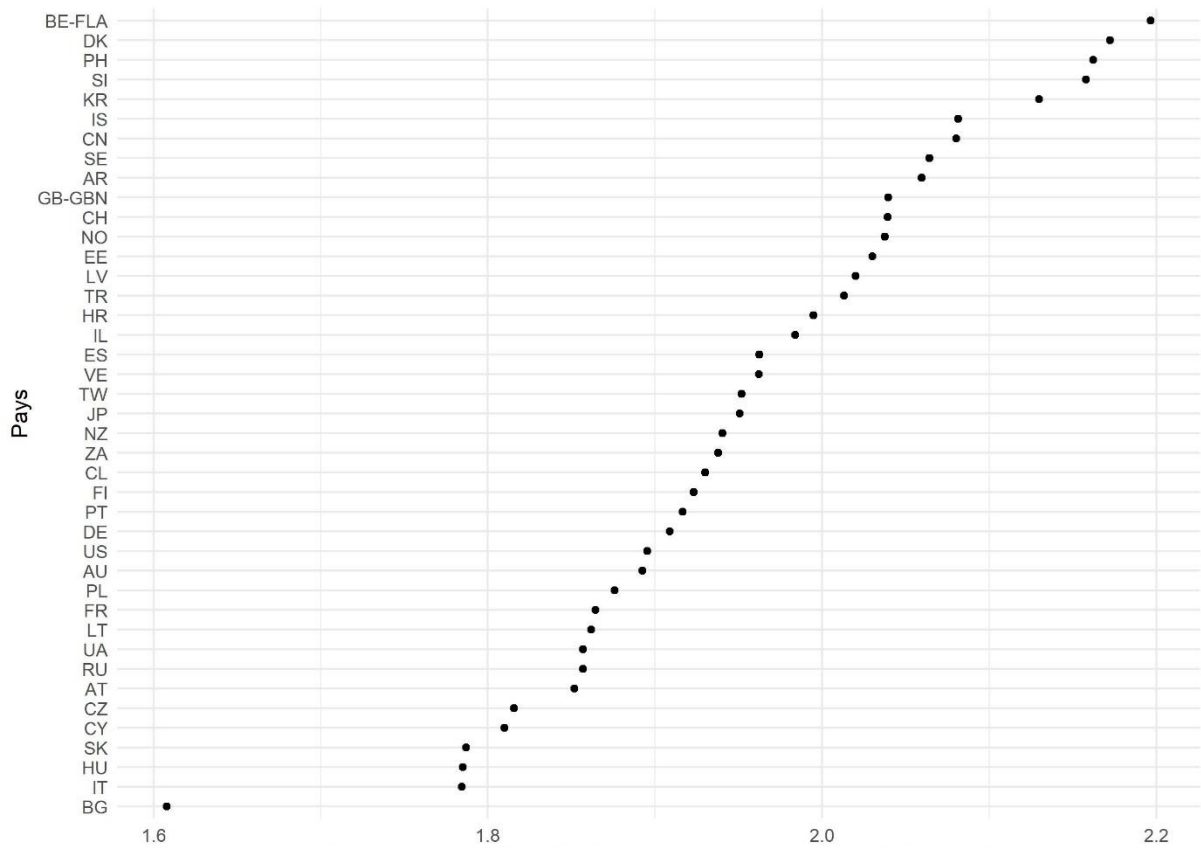


Fig. 5. Réponses à la question liée à l'importance des responsabilités dans le travail.

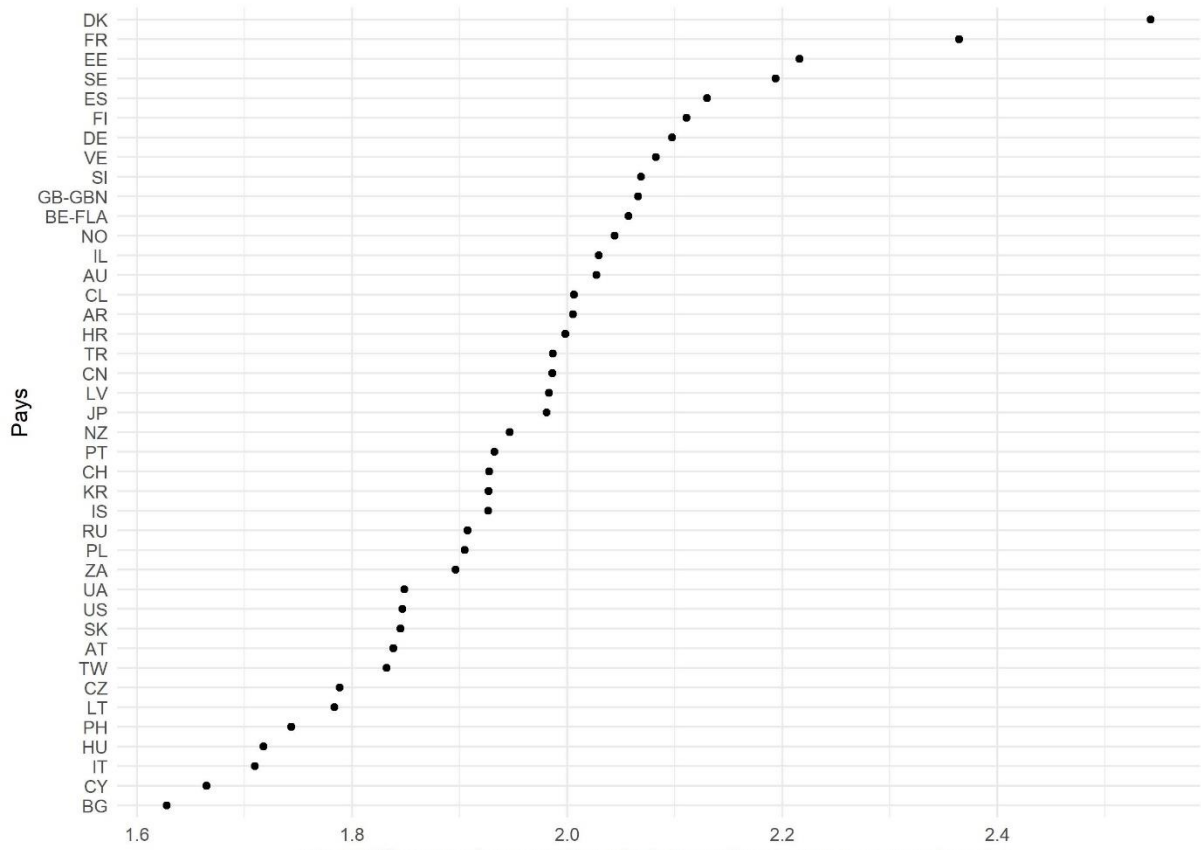


Fig. 6. Réponses à la question liée à l'importance du fait de travailler dur.

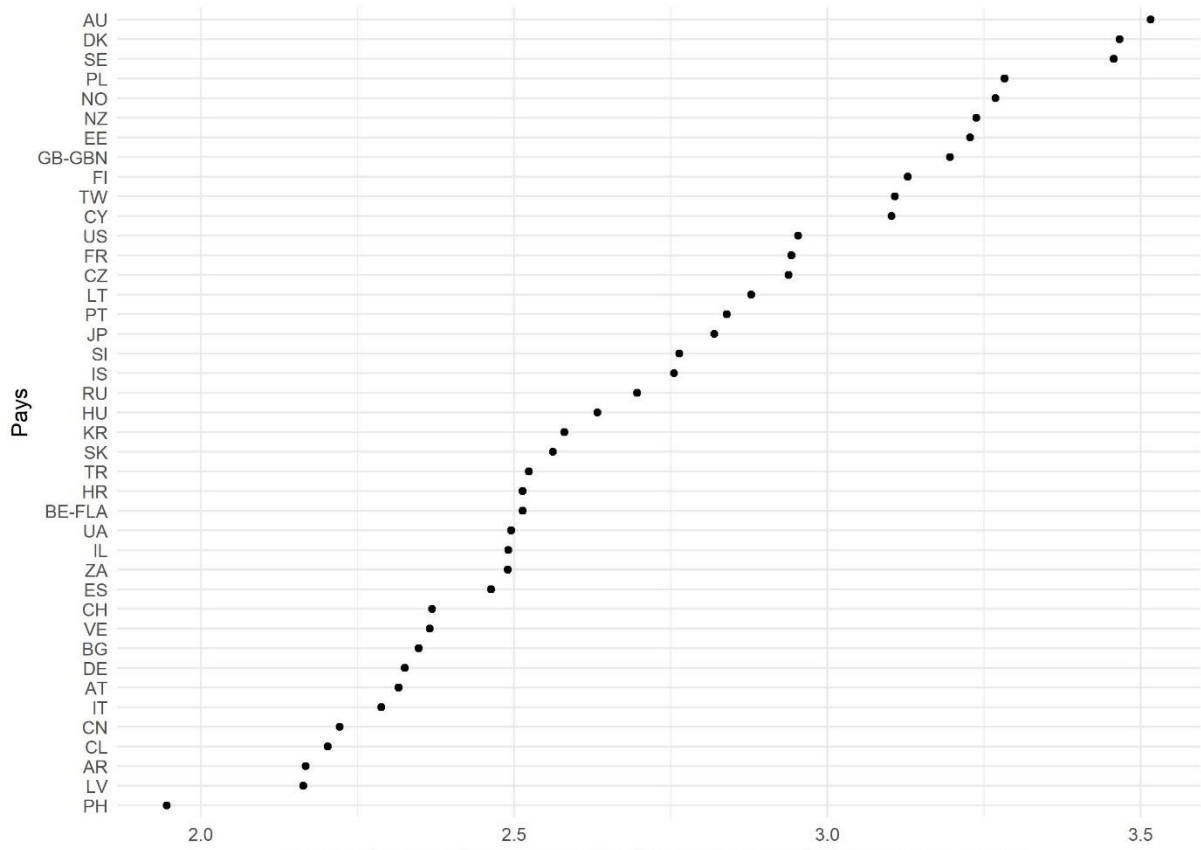


Fig. 7. Réponses à la question liée à l'importance du fait d'avoir des enfants.

Deuxième partie

De l'importance des données empiriques pour une
philosophie normative

Chapitre 4

Pourquoi une philosophie expérimentale ?

De l'usage des données empiriques pour informer le débat philosophique

Dans le cadre de cette thèse, nous mobiliserons des données empiriques, essentiellement issues de la psychologie sociale, pour informer les débats philosophiques relatifs au mérite. Une telle proposition met immédiatement en jeu un certain nombre de débats traditionnels portant sur la division propre entre philosophie et sciences, et sur le supposé gouffre séparant questions empiriques et questions normatives. Quelle est l'importance de données empiriques pour la philosophie morale et politique ? Est-il besoin d'acquérir une connaissance de la psychologie humaine pour construire une théorie normative ? En quoi la science peut-elle influencer les valeurs ?

Déterminer la relation entre science et valeur peut ouvrir le champ à un nombre très varié de positions, que l'on peut placer dans un continuum selon qu'elles accordent une plus ou moins grande importance à la science dans la détermination des questions normatives. A un extrême, on pourrait soutenir que la science peut apporter une réponse exacte à tout questionnement sur les valeurs morales, et qu'il est ainsi possible d'obtenir une connaissance à peu près certaine de ce qui doit être. Cette théorie intellectualiste de la morale a naturellement une longue histoire, et trouve son origine dans les conceptions platoniciennes qui considéraient qu'une science parfaite pouvait conduire à une parfaite moralité. Même si une telle théorie a plutôt perdu en popularité dans les débats philosophiques, on peut encore trouver des auteurs défendant une telle position, comme l'essayiste Sam Harris, auteur extrêmement populaire dans les pays anglo-saxons⁸⁹. A l'opposé, de nombreuses positions philosophiques classiques n'accordent

⁸⁹ Sam HARRIS, *Moral landscape: how science can determine human values*, New York, Free Press, 2011.

qu'une place limitée aux données empiriques, les réduisant à un rôle purement instrumental⁹⁰. Défendre l'idée que les données empiriques ne doivent jouer un rôle que purement instrumental, c'est considérer que, à partir du moment où une théorie morale est constituée, et une fois que les buts normatifs sont irrémédiablement fixés, il est nécessaire d'obtenir certaines connaissances empiriques pour savoir quels moyens sont les mieux à même de servir nos objectifs moraux. Une théorie sentimentaliste comme celle défendue par Hume correspond à ce schéma, de même qu'une théorie kantienne, qui dérive *a priori* l'impératif catégorique avant de faire appel à des données empiriques pour déduire les devoirs particuliers.

Une théorie radicalement scientifique, qui défend la possibilité d'une appréhension parfaite des valeurs par une recherche strictement scientifique, semble à première vue peu vraisemblable. Faute de croire en l'existence de valeurs dans un monde idéal auquel la science seule aurait accès, il est difficile de croire qu'une étude empirique puisse à elle seule décider de la manière dont nous devons vivre. Faut-il cependant se limiter à la croyance en un rôle purement instrumental des valeurs empiriques ?

Au-delà d'un rôle purement instrumental, il s'agira ici de défendre l'importance de données empiriques et psychologiques pour l'évaluation des théories morales elles-mêmes. Cette défense se fera en deux temps. Dans ce chapitre, j'établirai que toute philosophie morale repose fondamentalement sur des intuitions relatives à ce qu'il est moralement bon de faire, et qu'il est, dans ce cadre, essentiel de baser une théorie morale sur des intuitions courantes. Cette justification méta-éthique se fera à partir du rejet de toute théorie morale qui considérerait que des faits objectifs existent en éthique, indépendamment des opinions particulières des individus à leur sujet.

Après avoir démontré l'importance des intuitions morales en général, le chapitre suivant examinera la question de savoir dans quelle mesure les philosophes peuvent prétendre posséder des intuitions plus fines, ou plus exactes, que les individus ordinaires en matière de morale. Il s'agira donc de déterminer dans quelle mesure les philosophes peuvent prétendre avoir des intuitions morales moins biaisées que celles des individus ordinaires.

La période actuelle constitue un moment intense de réflexion sur l'utilité de données empiriques pour la philosophie morale et politique. Ce renouveau de la réflexion

⁹⁰ En théorie, il est bien sûr possible d'imaginer des théories n'accordant absolument aucune importance aux données empiriques vis-à-vis des problèmes normatifs. On peut par exemple imaginer qu'il est possible de déterminer absolument *a priori* tous les devoirs qu'a un individu, et que ces devoirs n'ont besoin d'aucun support empirique pour être réalisés. Par exemple, si l'on imagine une éthique défendant qu'il est possible de fournir des preuves *a priori* de l'existence de Dieu, et que le seul devoir humain est de prier, on pourrait considérer qu'une telle position morale n'a besoin d'aucun support empirique. De telles théories sont cependant si peu plausibles qu'elles sont naturellement absentes des discussions philosophiques.

méthodologique en théorie morale s'est déroulé sur trois fronts, assez largement indépendants. Tout d'abord, depuis les années 80, les progrès et le succès public de la sociobiologie, puis de la psychologie évolutionniste, ont conduit à une réflexion étendue sur l'impact de l'évolution sur les théories éthiques et méta-éthiques. Au cours des dernières années, des philosophes comme Richard Joyce et Sharon Street ont pu se servir de la théorie évolutionnaire pour opposer des arguments contre le réalisme moral⁹¹. Sur un autre front, depuis les années 2000 et le développement de la philosophie expérimentale, un débat important s'est développé autour de la question de l'importance des données psychologiques pour évaluer la validité d'une théorie philosophique. Ce débat a donné lieu à des évaluations très contrastées, certains philosophes défendant l'idée que la philosophie expérimentale est incapable d'apporter quoi que ce soit aux débats philosophiques, tandis que des partisans de cette approche pouvaient y voir une menace radicale pour la méthode philosophique classique faisant appel aux expériences de pensée⁹². Enfin, dans un mouvement indépendant, les années 2000 ont vu apparaître un « methodological turn » en philosophie politique. Un certain nombre de philosophes politiques ont défendu l'idée que les théories abstraites n'avaient en elles-mêmes aucune valeur tant qu'elles ne se fondaient pas sur les pratiques politiques concrètes des sociétés réelles. Ces nouvelles théories dites « réalistes », « non-idéales » ou « pratiques », se sont opposées aux théories considérées trop formelles développées dans la tradition analytique depuis Rawls, et ont défendu l'idée que le but de la philosophie était avant tout d'être la traduction des pratiques politiques effectives⁹³.

Justifications préliminaires : applications pratiques, importance d'une philosophie publique, rejet de l'utopisme.

Avant de nous lancer dans le vif du sujet, commençons par examiner trois justifications du recours aux données empiriques qui sont fondamentalement vraies, mais qui ne sont pas parfaitement satisfaisantes : 1) le recours instrumental aux données empiriques afin de tirer des

⁹¹ Richard JOYCE, *The evolution of morality*, Cambridge, Mass., The MIT Press, 2007 ; Sharon STREET, « A Darwinian Dilemma for Realist Theories of Value », *Philosophical Studies*, 2006, vol. 127, n° 1, p. 109-166.

⁹² Timothy WILLIAMSON, « Philosophical Expertise and the Burden of Proof », *Metaphilosophy*, 2011, vol. 42, n° 3, p. 215-229 ; Joshua ALEXANDER, Ronald MALLON et Jonathan M. WEINBERG, « Accentuate the Negative », *Review of Philosophy and Psychology*, 2010, vol. 1, n° 2, p. 297-314.

⁹³ Eva ERMAN et Niklas MÖLLER, « Practices and Principles: On the Methodological Turn in Political Theory », *Philosophy Compass*, 2015, vol. 10, n° 8, p. 533-546 ; David ENOCH, « Ideal Theory, Utopianism, and What's the Question (in Political Philosophy) ».

applications pratiques des théories philosophiques, 2) le recours aux données empiriques pour déterminer dans quelle mesure le philosophe doit chercher à conserver ou à réviser les idées populaires, et 3) le recours aux données empiriques visant à éviter le spectre d'une philosophie purement utopique.

Considérons brièvement la première raison. Tout d'abord, il est exact que des données empiriques permettent de savoir comment mettre en œuvre des théories philosophiques ; par exemple, si une philosophie normative indique qu'il est nécessaire de subvenir aux besoins de tous les concitoyens, il sera nécessaire d'obtenir des données économiques ou sociologiques pour savoir comment une telle théorie doit être mise en œuvre. De la même manière, il serait nécessaire d'obtenir des données empiriques sur les opinions des individus à propos de la justice sociale et de la satisfaction des besoins, pour savoir dans quelle mesure une politique de satisfaction des besoins se heurterait à une opposition, ou à un soutien massif, de la part de la population. Cette fonction instrumentale de l'obtention de données empiriques est évidente. Cependant, il ne s'agit pas d'une justification forte de l'utilité des données empiriques pour la philosophie. Si les données empiriques ne jouent qu'un rôle instrumental, alors la philosophie pourrait souverainement s'en passer pour la détermination des fins dernières. Ne pas se servir de données empiriques ne représenteraient en rien un coût pour l'élaboration d'une philosophie normative générale.

La deuxième justification de l'usage de données psychologiques correspond à l'idée que le philosophe doit connaître les opinions courantes relatives aux problèmes philosophiques afin de pouvoir communiquer avec le grand public. Si l'on adopte une vision pratique de l'activité philosophique, le philosophe doit être un philosophe public, qui cherche à adopter l'opinion correcte afin de pouvoir améliorer les conceptions défendues au sein de la société. Si le philosophe cherche à adopter la conception la plus cohérente ou la mieux défendable possible, ce doit être pour la proposer comme une conception pouvant être adoptée par l'ensemble de ses concitoyens. Il me semble que c'est, tout du moins, la conception de la philosophie la plus en phase avec le fait qu'une grande partie de philosophes professionnels sont financés par le public, qui doit au moins attendre des philosophes qu'ils puissent *a minima* justifier leur activité.

Dans ce cadre, si le philosophe a pour devoir d'entrer en activité de dialogue avec la société, il semble naturellement fondamental qu'il ait une connaissance exacte des opinions répandues au sein de la société. Autrement dit, si le philosophe dispose d'une théorie qu'il considère comme vraie, il doit se poser la question de savoir dans quelle mesure cette conception « vraie » est compatible avec les conceptions courantes. En fonction de cette compatibilité plusieurs

possibilités s'ouvrent, qui ont été bien mises en avant par Manuel Vargas dans ses travaux récents⁹⁴.

Vargas défend la distinction entre trois grands types d'approches relatives aux conceptions courantes : une théorie philosophique peut être conservatrice, éliminativiste, ou révisionniste. Une théorie est conservatrice si elle défend l'idée que les théories courantes sont fondamentalement correctes, et cherche donc à proposer une justification plus sophistiquée de conceptions qui sont déjà adoptées au sein de la population. Dans le cas du mérite, cela pourrait par exemple consister à dire que le mérite est une notion fondamentalement valable, et que les individus ont raison de lui accorder la place qu'elle possède. A l'inverse, une théorie est éliminativiste si elle considère que les conceptions courantes sont fondamentalement erronées, et qu'il serait préférable que les individus abandonnent complètement le concept en question. Il serait par exemple possible dans ce cas de défendre l'idée que le mérite est une conception empliée d'idées contradictoires, et fondamentalement indéfendable. Enfin, une théorie révisionniste est une théorie qui défend l'idée que les conceptions courantes sont globalement dans l'erreur, mais que le concept sert une fonction cruciale qui ne peut être simplement mise de côté. Au lieu de recommander l'élimination complète du concept, les théories révisionnistes recommandent alors sa transformation, afin de créer une conception qui remplisse la même fonction que le concept remplissait dans la vie courante, mais qui soit expurgée des contradictions et des problèmes qui étaient liés à l'utilisation naïve du concept. Dans le cas qui intéresse Vargas, celui-ci propose une conception révisionniste du libre-arbitre, qui évite les présupposés libertariens du concept, et limite l'idée d'une action libre à toute action qui peut faire l'objet d'un retour communicatif positif ou négatif. Dans le cas du mérite, comme nous l'avons vu dans le premier chapitre, la conception révisionniste la plus célèbre est celle que Rawls propose dans sa *Théorie de la justice*. Rawls défend en effet l'idée que mériter une récompense doit être interprété comme le fait d'avoir des attentes légitimes quant au fait que l'on va bien recevoir cette récompense. Il s'agit d'une reformulation qui est la plupart du temps en accord avec la conception naïve de ce que signifie mériter quelque chose, mais qui n'est manifestement pas une interprétation stricte de l'idée ordinaire de mérite.

Si un philosophe a déterminé quelle position était la plus cohérente et la mieux défendable vis-à-vis d'un problème philosophique particulier, alors il aura besoin de connaître les opinions ordinaires à ce sujet pour pouvoir déterminer dans quelle mesure il faut conserver, réviser ou

⁹⁴ Manuel VARGAS, « Revisionist Accounts of Free Will: Origins, Varieties, and Challenges », in Robert KANE (dir.), *Oxford Handbook on Free Will, 2nd Edition*, Oxford University Press, 2011, p. ; Manuel VARGAS, *Building Better Beings: A Theory of Moral Responsibility*, Oxford, Oxford University Press, 2015.

éliminer les opinions populaires. Nous pensons qu'il s'agit d'un rôle important de la philosophie expérimentale. La philosophie doit avoir un rôle public, et doit s'adresser à l'ensemble des citoyens. Pour cela, connaître leurs opinions au sujet de problèmes philosophiques est essentiel.

Cependant, si l'on adopte cette perspective tout en considérant que le philosophe doit atteindre la vérité indépendamment de l'opinion populaire, alors la fonction des données empiriques serait purement à visée communicative. La seule utilité de données empiriques serait de permettre au philosophe d'ajuster son discours public. Ce n'est pas rien, mais, en même temps, cela ne change pas fondamentalement la pratique philosophique habituelle. Comme nous allons le voir, nous pensons que les données empiriques doivent informer l'activité philosophique elle-même, et pas seulement sa communication au grand public.

Passons maintenant à la troisième justification : l'obtention de données expérimentales permettrait de rejeter des théories philosophiques trop utopiques. Je vais m'attarder un peu plus longuement sur cette troisième justification. Mais, pour le dire en un mot, il me semble que si le rejet de l'utopie justifie bien, dans une certaine mesure, le recours aux données empiriques, cette utilité est relativement limitée, étant donné qu'une quantité bien faible de philosophies politiques semblent tomber dans l'utopie pure et simple.

À première vue, on pourrait avoir l'impression que la fonction la plus forte de l'usage de données empiriques pourrait être d'éviter le danger de l'élaboration d'une philosophie purement utopique. Une telle volonté de rejet de l'utopie n'est pas en soi absurde ; il est en effet légitime de considérer qu'une philosophie politique qu'il serait rigoureusement impossible d'appliquer, ou d'approcher d'une quelconque manière, n'aurait qu'un intérêt très limité. Se raccrocher au possible semble indispensable pour élaborer une théorie qui possède un intérêt minimal. Et, dans ce cadre, la considération de données empiriques semble fondamentale pour élaborer une théorie politique qui évite le risque de l'utopie.

Il est important, toutefois, de ne pas aller trop loin dans ce rejet de politiques improbables, c'est-à-dire dans ce que David Estlund a critiqué sous le nom d'utopophobia⁹⁵. Il est en effet très difficile de tracer la frontière entre ce qui relèverait d'une politique utopique, et ce qui est simplement une politique ambitieuse, mais parfaitement réalisable. De fait, sauf à se complaire dans l'adhésion pure et simple à l'ordre existant, toute politique un tant soit peu ambitieuse court un risque très fort de ne jamais être mise en place. Ce qui est problématique n'est donc pas qu'une politique ait une probabilité très faible d'être parfaitement réalisée. C'est, fondamentalement, le destin de toute politique. Ce qui est profondément néfaste dans l'idée

⁹⁵ David M. ESTLUND, *Democratic authority: a philosophical framework*, Princeton, NJ, Princeton Univ. Press, 2009.

d'utopie, c'est qu'il s'agit d'une philosophie politique qui est tellement coupée de la réalité qu'elle perd toute possibilité de guider l'action d'une quelconque manière.

Pour comprendre les limites et l'utilité d'une critique des théories dites « utopiques », il peut être utile de distinguer deux manières à partir desquelles une théorie peut être décrite comme utopique. Premièrement, une théorie peut être considérée comme utopique parce qu'elle a très peu de chances d'être mise en place. En tant que telle, cette description n'est pas fondamentalement problématique. En effet, une théorie qui ne sera jamais mise en place peut garder une réelle valeur pratique en tant qu'elle peut toujours servir d'idéal régulateur. Une politique peut en effet être connectée à la réalité de plusieurs manières : 1) parce qu'il est possible de la mettre en place dans son intégralité (réalisation intégrale), 2) parce qu'il est possible de la mettre en place de façon partielle (réalisation partielle), 3) parce qu'elle peut servir à évaluer des politiques ou situations alternatives (évaluation comparative). Tant qu'au moins une des trois connections est préservée, il semble clair que le manque de probabilité de réalisation d'une théorie ne sera jamais en soi un clair défaut. Même dans le cas d'une politique trop ambitieuse, il est toujours possible d'adopter la position qu'Estlund nomme un réalisme désespéré⁹⁶ : on peut continuer de défendre une politique, même en sachant qu'elle ne sera jamais mise en place.

Cependant, il faut bien admettre que désigner une politique comme utopique peut être une critique valable, mais cela requiert d'adopter un deuxième sens du terme « utopique ». Une politique peut être désignée comme utopique si elle fait des hypothèses sur la nature humaine ou sur des faits de la vie sociale qui sont en rupture avec les données effectives. Par exemple, si une théorie fait l'hypothèse que les individus sont toujours généreux ou se préoccupent uniquement des intérêts des plus faibles ou des plus démunis, alors une théorie pourra être décrite comme utopique. Et une telle théorie sera effectivement dénuée d'intérêt parce que, au lieu de s'appliquer à l'univers politique qui est le nôtre, elle ne s'appliquera qu'à un univers parallèle dont nous n'avons guère à nous soucier.

Ceci étant dit, combien de théories politiques majeures ont fait des hypothèses farfelues sur la nature humaine ? Fondamentalement, très peu. La plupart des grandes théories politiques modernes, qu'elles soient utilitaristes ou kantienne, admettent le côté intéressé et parfois égoïste des individus. Elles présupposent, en général, l'existence d'un sens de la justice, mais sans jamais faire l'hypothèse que ce sens de la justice soit tout puissant. Il s'agit, en général, de théories qui se basent sur des présupposés généraux très faibles sur la nature humaine,

⁹⁶ *Ibid.*

présupposés qui ont peu de chances d'être rejetés par de nouvelles découvertes empiriques. En conséquence, étant donné que l'utopie est un risque à peu près inexistant, le danger de l'utopie ne peut servir de justification principale pour l'élaboration d'une philosophie informée empiriquement.

Nous avons donc déterminé trois fonctions premières de l'obtention de données empiriques : il s'agit d'obtenir des informations quant aux moyens d'appliquer les théories philosophiques, de savoir s'il faut réviser les opinions populaires, et il s'agit de nous garder des théories bâties sur des données fondamentalement incompatibles avec la nature humaine.

Nous avons établi que ces trois fonctions ne sont pas en elles-mêmes décisives, en tant qu'elles semblent exclure un nombre très faible de théories normatives. Dans le cas de l'évitement du danger de l'utopisme, par exemple, il n'est pas facile de dire où se trouvent les hypothèses irréalistes quant à la nature humaine dans la *Théorie de la justice* ou dans la plupart dans des grandes théories normatives modernes⁹⁷. La plupart de ces théories sont en effet des théories ambitieuses, mais sans que cette ambition dans la définition de ce que serait une société juste implique une négation de la réalité, ou implique de mentir au sujet de la faible probabilité de réalisation d'une société juste.

En conclusion, acquérir un sens des données empiriques pour éviter une fuite loin de la réalité peut être utile, mais n'est pas en soi suffisant pour justifier pleinement une philosophie expérimentale. Il est temps maintenant de nous tourner vers une justification pleine et entière de l'utilité des données expérimentales pour l'activité philosophique.

La vraie justification d'une philosophie expérimentale : l'absence d'existence autonome des idées morales

Nous allons maintenant examiner une des questions les plus importantes pour déterminer l'utilité des données empiriques pour construire une théorie philosophique. Cette question cruciale est : dans quelle mesure les concepts philosophiques possèdent-ils une réalité en dehors de l'esprit ? Et si l'on admet qu'il n'existe pas de réalité des conceptions morales en dehors des

⁹⁷ Naturellement, Rawls et d'autres auteurs proposent des expériences de pensée, comme le voile d'ignorance, qui ne reflètent pas les caractéristiques effectives de la vie réelle. Mais les accuser d'utopisme pour cette raison est absurde : une expérience de pensée est un modèle simplifié, qui vise à mettre uniquement en avant des traits saillants de la vie réelle afin d'aiguiser les intuitions normatives du philosophe. Personne ne prétend qu'une expérience de pensée est un reflet fidèle de la vie réelle.

esprits humains, et s'il n'existe pas de vérité qui puisse être définie comme correspondance avec ces réalités extérieures, quel peut être le rôle du philosophe ?

Pour motiver l'importance de ces questions, il est possible de reprendre une distinction opérée par les philosophes américains Shaun Nichols et Joshua Knobe dans leur manifeste pour la philosophie expérimentale⁹⁸. Les auteurs effectuent une distinction entre l'utilité d'enquêtes psychologiques pour la recherche en sciences naturelles, d'un côté, et pour la recherche en philosophie, de l'autre. Nichols et Knobe commencent par admettre qu'obtenir une connaissance des opinions populaires relatives à la physique ou à la biologie n'est évidemment que de peu d'intérêt pour comprendre la réalité des données physiques ou biologiques. Cependant, ils défendent l'idée que le cas de la philosophie est très différent : alors qu'il existe, sauf à défendre un idéaliste extrême, un critère extérieur objectif dans le cas des recherches en physique ou en biologie, un tel critère extérieur fait complètement défaut dans le cas de la plupart des recherches philosophiques. Par exemple, dans le cas des idées morales, ou des concepts avec une portée normative comme la connaissance, il n'est en rien évident de savoir en quoi consisterait la moralité, ou la connaissance, en dehors de l'existence d'individus ayant des opinions à leur sujet. Dans ces cas-là, il semble bien que les concepts vis-à-vis desquels nous souhaiterions obtenir une connaissance sont en bonne partie constitués par l'activité humaine. Déterminer l'importance des données psychologiques pour l'activité philosophique impose donc de s'interroger sur la réalité, ou les conditions d'existence, des objets d'étude philosophique.

Les différentes positions possibles en méta-éthique, et leurs implications

A partir de maintenant, nous nous concentrerons sur le mode d'existence des idées *morales*, étant donné qu'il s'agit de l'objet essentiel de cette thèse. En particulier, le lecteur pourra garder en tête la condition d'existence des idées de mérite, de justice, ou de moralité en général. Je vais brièvement esquisser les différentes possibilités théoriques quant au mode d'existence de ces concepts, avant d'indiquer les conséquences possibles pour une philosophie expérimentale.

⁹⁸ Joshua KNOBE et Shaun NICHOLS, « An Experimental Philosophy Manifesto », in Joshua KNOBE et Shaun NICHOLS (dir.), *Experimental Philosophy*, Oxford University Press, 2007, p. 3–14.

Deux grandes options sont possibles en ce qui concerne le mode d'existence des idées morales. Premièrement, il est possible d'adopter une position dite réaliste, et considérer que les objets philosophiques correspondent à des objets existant indépendamment des esprits humains, à la manière dont un atome existe indépendamment d'un humain qui l'observe. Il existerait alors des objets extérieurs correspondant à nos idées de mérite, de justice, ou de moralité. À l'inverse, il est possible d'adopter une position méta-éthique non-réaliste, et de considérer que les concepts moraux sont constitués par les idées que les individus se font à leur sujet, et dépendent donc des détails de la psychologie humaine.

Que l'on adopte une perspective réaliste ou non-réaliste, il est possible d'adopter une position favorable ou hostile à l'utilité d'une étude empirique des conceptions ordinaires. Dans le cas d'une philosophie réaliste, on peut en effet adopter deux positions différentes : ou bien on considère que les idées morales populaires reflètent les réalités morales indépendantes, ou bien l'on considère que les idées morales quotidiennes sont sans rapport avec les réalités morales. Dans le premier cas, la psychologie morale peut apporter des indices de l'existence d'objets moraux, et possède donc une grande pertinence. Dans le second cas, la nature des objets moraux doit être déterminée indépendamment des idées courantes à leur sujet, et la psychologie morale ne présente aucun intérêt particulier⁹⁹.

En ce qui concerne les perspectives non-réalistes, celles-ci peuvent à nouveau se diviser en deux : on peut être ou bien constructiviste, et considérer qu'il existe une vérité morale, même si cette vérité est dépendante de la psychologie des individus concernés, ou bien considérer qu'il n'existe aucune vérité en matière morale, et adopter ainsi une perspective que l'on pourrait qualifier de nihiliste. Dans ce dernier cas, il serait douteux que l'on puisse construire une éthique normative. Mais même dans le premier cas, si l'on adopte une perspective constructiviste, il n'est toujours pas évident que la philosophie expérimentale présente un intérêt. Il est en effet possible d'adopter un constructivisme kantien ou universaliste, qui considère que les vérités morales s'imposent à toute créature rationnelle, même si ces vérités morales ne possèdent pas d'existence séparée en tant qu'objet. Si l'on adopte une telle perspective universaliste, il est douteux qu'obtenir des données sur les opinions des gens ait de l'importance pour une philosophie morale normative, étant donné que les règles sont supposées être universelles et connaissables *a priori*. C'est donc uniquement dans le dernier cas, celui

⁹⁹ Je remercie Bertrand Guillaume pour m'avoir poussé à préciser les différentes options possibles s'offrant à une perspective méta-éthique réaliste.

d'un constructivisme qu'on pourrait qualifier de humien à la suite de Sharon Street¹⁰⁰, que les données expérimentales prennent toute leur importance : si l'on considère que des vérités morales existent, mais qu'elles dépendent de la psychologie et des croyances précises d'individus particuliers, alors il devient essentiel d'obtenir des données sur les intuitions morales des individus¹⁰¹.

Je vais revenir sur ces différentes perspectives possibles au cours de ce chapitre. Dans l'immédiat, notons qu'il existe deux positions méta-éthiques possibles accordant clairement de l'importance aux données empiriques : une perspective réaliste qui considère que les idées morales ordinaires sont le reflet d'objets moraux existant indépendamment de la pensée humaine, et une perspective constructiviste humienne qui défend l'idée que les vérités morales dépendent de la constitution psychologique particulière des individus. Les trois autres perspectives que nous avons identifiées (réaliste refusant le lien entre vérités morales et idées courantes, constructivisme universaliste, nihilisme) ne préconiseraient probablement pas le recours aux idées ordinaires pour la construction d'une philosophie normative.

Mon argument procédera en deux temps. Dans un premier temps, je montrerai que des perspectives méta-éthiques non-réalistes sont beaucoup plus plausibles que des perspectives réalistes. Dans un deuxième temps, je montrerai qu'un constructivisme humien est préférable à un perspectivisme kantien ou à une position nihiliste. Étant donné que le constructivisme humien accorde de l'importance aux données psychologiques, j'aurai ainsi démontré la pertinence d'une approche expérimentale en philosophie.

Avant de continuer, il me faut bien sûr confesser que je n'ai pas les moyens de résoudre toutes ces questions méta-philosophiques ou méta-éthiques de façon parfaitement satisfaisante. Cette question est cependant si importante pour mon propos que je présenterai les arguments qui me semblent les plus pertinents pour indiquer pourquoi je considère qu'un constructivisme humien est la position la plus plausible en méta-éthique. Pour ce faire, je reprendrai les principaux arguments développés par Sharon Street pour étayer son rejet du réalisme moral et pour favoriser sa théorie du constructivisme humien.

¹⁰⁰ Sharon STREET, « Coming to Terms with Contingency : Humean Constructivism About Practical Reason », in Jimmy LENMAN et Yonatan SHEMMER (dir.), *Constructivism in Practical Philosophy*, Oxford University Press, 2012.

¹⁰¹ Une perspective culturaliste relativiste peut être considérée comme un cas particulier du constructivisme humien, en tant que le relativisme culturel admet qu'il existe des vérités morales, mais que celles-ci sont relatives à la culture dominante au sein de différentes sociétés. Une telle perspective considérerait évidemment que des données empiriques, de préférence inter-culturelles, sont essentielles pour construire une philosophie morale normative.

Contre le réalisme moral en méta-éthique

Avant de développer les arguments avancés par Street contre le réalisme moral, il convient de remarquer que l'idée d'une existence indépendante des idées morales semble extrêmement coûteuse d'un point de vue ontologique. Il est évident que, si les idées morales comme la justice ou le mérite possèdent une existence en dehors de l'esprit, ce type d'existence ne saurait être celui commun aux objets physiques ordinaires. Il s'agirait donc d'un type d'existence particulier, ce qui est loin d'être parcimonieux d'un point de vue ontologique.

De plus, le réaliste moral aurait l'obligation de préciser de quelle manière nous pourrions avoir accès à ces idées morales, et quel mode de connaissance serait alors permis¹⁰². Bien sûr, de nombreuses théories en méta-éthique ont défendu qu'une connaissance d'objets moraux existant indépendamment de l'esprit est possible¹⁰³. Cependant, il est important de noter que ces théories partent avec un désavantage, en tant qu'elles doivent apporter des arguments qui permettent de surmonter le manque d'économie ontologique.

Au-delà de l'argument général de simplicité, Sharon Street, dans un article récent, a montré comment la théorie de l'évolution peut être mobilisée pour démontrer le manque de vraisemblance des théories réalistes en méta-éthique¹⁰⁴. L'article de Sharon Street se fonde sur l'idée que les comportements et les idées morales ont été fondamentalement influencés par l'évolution humaine, et donc qu'elles reflètent des pressions évolutives qui se sont exercées dans l'environnement ancestral humain¹⁰⁵. Sharon Street elle-même cite plusieurs exemples particulièrement frappants de comportements qui sont un indice de l'impact de l'évolution sur le façonnement du raisonnement humain : « (1) Le fait que quelque chose promouvrait notre survie est une raison en sa faveur. (2) Le fait que quelque chose servirait les intérêts d'un membre de notre famille est une raison en sa faveur. (3) Nous avons de plus grandes obligations

¹⁰² John L. MACKIE, *Ethics: inventing right and wrong*, London, Penguin Books, 1990.

¹⁰³ Pour certains travaux récents, voir Russ SHAFER-LANDAU, *Moral realism: a defence*, Oxford, Clarendon Press, 2005 ; David ENOCH, *Taking morality seriously: a defense of robust realism*, Oxford, Oxford Univ. Press, 2013.

¹⁰⁴ S. STREET, « A Darwinian Dilemma for Realist Theories of Value », *op. cit.*

¹⁰⁵ Même si une telle perspective est rarement adoptée en sciences sociales en France, elle est en elle-même difficilement réfutable, et a fait l'objet d'une démonstration intensive au cours des programmes de recherche en psychologie évolutionniste au cours des quarante dernières années. Si jamais le lecteur n'est pas convaincu par la viabilité d'une perspective évolutionniste pour expliquer la moralité humaine, il pourra simplement voir les paragraphes suivants comme décrivant un argument hypothétique, valide uniquement en fonction d'un éventuel progrès futur en psychologie évolutionniste. Pour simplifier l'exposition ici-même, j'ai résumé les principaux arguments en faveur d'une approche évolutionniste dans le cadre de l'appendice au présent chapitre. Les lignes qui suivent se fondent donc sur l'hypothèse que l'évolution a eu une influence déterminante sur la pensée morale humaine.

d'aider nos propres enfants que nous n'en avons d'aider de parfaits étrangers. (4) Le fait que quelqu'un m'a bien traitée est une raison de bien traiter cette personne en retour. (5) Le fait que quelqu'un est altruiste est une raison de l'admirer, de la louer, et de la récompenser. (6) Le fait que quelqu'un a causé un tort délibéré est une raison de l'éviter ou de chercher à la punir¹⁰⁶. » Tous ces traits sont très largement partagés d'un point de vue inter-culturel, et il est relativement évident qu'ils favorisent une plus grande survie et descendance des individus qui les adoptent. Cette perspective est davantage développée dans l'appendice à ce chapitre. Dans l'immédiat, je demanderai simplement au lecteur de faire l'hypothèse que les idées morales humaines ont bien été façonnées par des forces évolutionnistes.

Si l'on admet que les idées morales dépendent de processus évolutifs, quelle conséquence peut-on tirer sur la validité du réalisme moral ? Street considère que la théorie évolutionniste pose un dilemme au réalisme moral. Le réaliste postule qu'il existe des idées morales indépendantes de la pensée humaine. La plupart des réalistes postulent également que les idées morales ordinaires correspondent globalement aux idées morales objectivement valides, et existantes indépendamment de l'esprit. Le problème évolutionniste qui se pose au réaliste est de déterminer comment une telle correspondance est possible, sachant qu'il existe une explication indépendante de l'apparition des idées morales. En un mot, Sharon Street défend l'idée que nous disposons d'une très bonne explication pour l'apparition des idées morales, à savoir la théorie évolutionniste. Dans ce cadre, il devient superflu de faire appel à des objets existant indépendamment de l'esprit humain pour expliquer la moralité humaine. L'hypothèse de l'existence d'objets moraux indépendants devient parfaitement gratuite.

Selon Street, l'existence d'explications évolutionnistes conduit le réaliste à être confronté à deux possibilités : ou bien le réaliste considère que les idées morales n'ont pas exercé de pression lors du processus évolutif, ou bien il considère que l'adoption d'idées morales objectivement existantes a constitué un avantage évolutif qui a amélioré la survie des individus. Or, chacune de ces deux options pose de lourds problèmes pour une perspective réaliste.

Selon la première branche du dilemme, si le réaliste défend l'idée que les idées morales objectives n'ont pas eu d'influence sur l'évolution humaine, alors il serait bien en peine d'expliquer par quelle coïncidence les idées morales que nous possédons, et qui ont été produites par l'évolution, se trouveraient être les mêmes que les idées morales objectives. En l'absence de raison expliquant pourquoi les deux ensembles d'idées morales se trouveraient

¹⁰⁶ S. STREET, « A Darwinian Dilemma for Realist Theories of Value », *op. cit.*, p. 115.

être les mêmes, le réaliste se verrait défendre une théorie ayant perdu encore davantage en plausibilité.

Selon la deuxième branche du dilemme, si le réaliste défend l'idée que percevoir les idées morales objectives a offert un avantage en termes de survie et de reproduction, alors il se trouve proposer une théorie nouvelle en biologie évolutionniste : avoir une connaissance d'idées morales existant indépendamment de l'esprit offrirait un avantage évolutif. A première vue, une telle théorie semble plausible : les idées morales (ou certaines d'entre elles du moins) offrent un avantage pour favoriser la transmission génétique. Cependant, un tel avantage peut se faire même en l'absence de l'existence indépendante d'idées morales. Et si les mêmes phénomènes peuvent s'expliquer sans postuler l'existence d'idées morales objectives, il est alors plus parcimonieux, d'un point de vue scientifique, de refuser l'idée que percevoir des idées morales objectives ait eu un quelconque avantage adaptatif, par opposition au simple fait de créer des idées morales de toute pièce. En conséquence, dans tous les cas, un réaliste serait bien incapable d'expliquer pourquoi il existerait un lien entre idées morales humaines et vérités morales objectives. Une hypothèse méta-éthique non réaliste, qu'elle soit constructiviste ou relativiste, semblerait alors beaucoup plus plausible.

Bien sûr, ces arguments ne prouvent pas la fausseté du réalisme en méta-éthique. Mais ils montrent qu'il s'agit d'une position philosophique partant avec un fort désavantage en termes de plausibilité et de parcimonie. Si le lecteur m'a suivi sur ce point, il convient maintenant de se rabattre sur l'autre grande option en méta-éthique permettant de justifier l'élaboration d'une théorie normative : le constructivisme.

De l'importance des données empiriques dans le cadre d'une philosophie morale non réaliste : la place du constructivisme

Si l'on considère qu'il n'existe pas de réalités morales indépendantes de la psychologie des individus, alors il convient de se tourner vers une autre théorie en méta-éthique, le constructivisme. Comme nous l'avons vu, celui-ci peut être associé à deux positions : une position universaliste et une position non-universaliste. Pour comprendre l'importance des données empiriques dans ce cadre, je devrai ici brièvement rappeler les débats contemporains autour du constructivisme en méta-éthique.

Selon la définition donnée par l'article de la *Stanford Encyclopedia*, « Le constructivisme méta-éthique correspond à la théorie selon laquelle, dans la mesure où il existe des vérités normatives, elles ne sont pas fixées par des faits normatifs qui sont indépendants de ce sur quoi des agents rationnels se mettraient d'accord selon des conditions définies de choix¹⁰⁷. » Le constructivisme cherche donc à savoir dans quelle mesure il est encore possible de découvrir l'existence de vérités morales, même dans un contexte où l'on ne reconnaît pas de vérités morales extérieures aux choix des individus. Il s'agit donc de savoir dans quelle mesure il est possible de trouver un moyen de découvrir une forme d'objectivité en morale, même si cette objectivité se trouve être relative aux valeurs et à la perspective d'agents rationnels.

Dans les débats contemporains sur le constructivisme, deux formes principales sont apparues : le constructivisme dit kantien, défendu notamment par la philosophe Christine Korsgaard, et le constructivisme dit humien, défendu notamment par Sharon Street¹⁰⁸. La principale différence entre ces deux formes de constructivisme porte sur la question de savoir dans quelle mesure des standards universels de normativité peuvent être développés, ou si les standards obtenus seront toujours relatifs aux valeurs et aux normes portées par les individus, sans qu'il soit possible de garantir une convergence universelle entre les valeurs des différents individus. La perspective universelle est propre au constructivisme kantien, la perspective relative est propre au constructivisme humien.

Comme nous l'avons vu, si l'on considère qu'il existe des raisons morales qui s'imposent à tous les êtres rationnels, comme le soutient le constructivisme kantien, alors il devient improbable qu'une enquête empirique ait beaucoup de valeur philosophique. S'il existait une méthode *a priori* permettant de découvrir une universalité des jugements moraux non seulement entre êtres humains, mais également vis-à-vis de toute espèce rationnelle possible, alors la connaissance d'idées morales à un endroit déterminé serait hautement superflue. Pour cette raison, je vais procéder à un examen du constructivisme kantien, avant de démontrer pourquoi une telle perspective est fondamentalement très improbable. Ceci me permettra, à la fin de ce

¹⁰⁷ Carla BAGNOLI, « Constructivism in Metaethics », in Edward N. ZALTA (dir.), *The Stanford Encyclopedia of Philosophy*, Winter 2017., Metaphysics Research Lab, Stanford University, 2017.

¹⁰⁸ Christine M. KORSGAARD, *The Sources of Normativity*, Cambridge ; New York, Cambridge University Press, 1996 ; S. STREET, « A Darwinian Dilemma for Realist Theories of Value », *op. cit.* ; John Rawls et Thomas Scanlon sont les deux auteurs les plus célèbres associés au constructivisme kantien ; cependant, il n'est pas tout à fait clair que leur position consiste réellement à adopter une théorie méta-éthique. Je reviendrai plus bas sur cette question dans le cas de Rawls. En ce qui concerne la position dite du constructivisme humien, elle est très proche des idées défendues notamment par Allan Gibbard. J'adopte ici la perspective de Sharon Street pour simplifier la discussion, et parce qu'il s'agit d'une perspective que je considère particulièrement bien développée. Cf. Allan GIBBARD, *Wise choices, apt feelings: a theory of normative judgment*, Oxford, Clarendon Press, 1990.

chapitre, de défendre la perspective d'un constructivisme humien et de mettre en avant l'utilité des données empiriques pour une philosophie normative.

Dans le cas du constructivisme kantien, Christine Korsgaard défend l'idée que le simple fait d'être un agent rationnel, portant un certain nombre de valeurs, implique que l'agent doit adhérer à certaines normes morales. En ce sens, selon Korsgaard, tout individu rationnel, qu'il soit humain ou non, quelles que soient sa culture ou ses idées personnelles, doit adhérer au même ensemble de normes et de jugements moraux. Il s'agit donc d'une exigence extrêmement forte. Le constructivisme kantien essaie de tenir ensemble deux idées qui peuvent sembler complètement opposées : d'une part, l'idée que les valeurs morales n'existent pas de façon indépendante, et doivent être rapportées à l'attitude pratique et à la nature des individus ; d'autre part, l'idée qu'il existerait des valeurs universelles, valables pour tout être raisonnable. Pour un constructiviste kantien, il s'agit donc de chercher un facteur propre aux êtres rationnels qui les conduiraient à devoir adopter certains jugements moraux.

Il est bien sûr difficile de considérer par avance tout type de constructivisme universaliste. Je préfère donc me concentrer sur le cas du constructivisme universaliste le plus célèbre, celui de Korsgaard, avant de montrer pourquoi, en général, une approche de ce type se heurte à des problèmes insurmontables. Je vais donc brièvement résumer ici la théorie de Korsgaard, montrer pourquoi je ne la considère pas comme une théorie plausible, avant de montrer en quoi une perspective plus modeste, comme le constructivisme humien, permet de maintenir un véritable sens à l'activité philosophique.

L'argument que Korsgaard développe dans *The Sources of Normativity* se fonde sur l'idée qu'un être rationnel doit valoriser des choses pour des raisons, et que ces raisons doivent se fonder à leur tour sur une base qui leur donne du sens¹⁰⁹. Si je considère qu'il est bon de transmettre certaines attitudes traditionnelles, par exemple le fait de valoriser la cuisine française, je dois le faire pour une certaine raison. Or, pour Korsgaard, si ces raisons peuvent être justifiées, elles doivent l'être par l'appel à une certaine identité portée par l'agent. Si je valorise la cuisine traditionnelle, c'est que cette attitude est justifiée par mon identité française. De même, si je valorise les études, cette attitude est justifiée par mon identité d'étudiant, ou bien, plus généralement, par mon identité d'individu connaissant. De manière générale, Korsgaard fait le présupposé que toute valeur doit reposer sur une certaine attitude réflexive, attitude réflexive qui à son tour implique la mise en avant d'une certaine identité. C'est parce que nous possédons une certaine identité que nous sommes capables de donner de la valeur à

¹⁰⁹ C. M. KORSGAARD, *The Sources of Normativity*, *op. cit.*

nos activités. Si je n'étais pas français, ou si je ne possédais pas l'identité d'un individu gourmet, je serais incapable de valoriser la cuisine traditionnelle.

Cette première étape dans le raisonnement de Korsgaard est en soi évidemment assez problématique. Il n'est en rien évident que l'identité individuelle soit le fondement nécessaire de toute valeur. Passons outre cependant. Dans une deuxième étape, Korsgaard part d'un deuxième présupposé : selon elle, les identités particulières que nous possédons doivent toutes reposer sur une identité fondamentale, à savoir, l'identité d'être raisonnable, ou d'être humain. Si je possède l'identité de Français, je suis obligé de posséder également l'identité d'être humain raisonnable. Selon Korsgaard, toute identité doit forcément se fonder sur l'identité plus fondamentale d'être humain capable d'adopter telle ou telle identité particulière. Dans le cas des communautés, l'identité française est un cas particulier de l'identité d'être humain capable d'être membre d'une certaine communauté culturelle.

Dans une troisième étape, Korsgaard défend l'idée que le fait d'adopter une certaine identité implique nécessairement de valoriser cette identité. Si mon identité est française, et si cette identité me conduit à valoriser certaines activités ou pratiques, alors je valorise nécessairement mon identité de Français. Si j'adopte une identité d'être humain, alors, nécessairement, je valorise mon humanité.

Enfin, dans une quatrième étape, Korsgaard conclut que le fait que je valorise mon humanité implique que je dois me reconnaître des obligations vis-à-vis de tout être humain.

Une réfutation complète et méthodique de tous ces arguments nous emmènerait trop loin¹¹⁰. Cependant, on peut dès à présent remarquer que chacune des étapes du raisonnement pose de nombreux problèmes. En ce qui concerne le premier argument, comme dit précédemment, il n'y a rien d'évident à ce qu'il soit impossible de valoriser quelque chose si l'on ne se fonde pas sur une identité particulière. Considérer que je ne peux pas valoriser « naïvement » le fait de faire du sport, d'aller à un concert, ou de maintenir des relations sociales, mais que toute valorisation de ce type doit se fonder sur une identité semble à première vue très peu plausible. En ce qui concerne la deuxième étape, on peut sans doute reconnaître que toute identité portée par un être humain présuppose une identité humaine. Cependant, il n'y a en soi rien d'évident à ce que posséder une identité conduise à valoriser cette identité. Enfin, il n'est pas forcément

¹¹⁰ Le lecteur intéressé pourra consulter notamment les critiques convaincantes de G. A. Cohen dans ses réponses à Korsgaard. G. A. COHEN, « Reason, humanity, and the moral law », in Onora O'NEILL (dir.), *The Sources of Normativity*, Cambridge ; New York, Cambridge University Press, 1996, p. 167-188.

très plausible de penser que valoriser sa propre humanité conduise à reconnaître des obligations vis-à-vis de toute l'humanité¹¹¹.

Au-delà du manque de plausibilité de la théorie particulière défendue par Korsgaard, il faut remarquer que toute théorie constructiviste universaliste présuppose de reconnaître des traits valables *a priori* pour tout être raisonnable possible et imaginable, et de défendre que ces traits constitutifs de la rationalité impliquent nécessairement une certaine attitude morale. La théorie morale défendue par Korsgaard nous permet de mettre en avant les failles de toute perspective universaliste, qui tenterait de fonder la morale dans une attitude pratique portée par tous les sujets. Pour justifier les attitudes morales ordinaires et leur contenu universel, Korsgaard en est amenée à faire appel à un formalisme vide, qui doit être capable de convenir à tous les êtres existants. Mais découvrir un ensemble de traits liés à toute attitude rationnelle possible et imaginable semble extrêmement délicat. Dans le cas de l'humanité, on peut évidemment soutenir que nous partageons tous un ensemble de gènes extrêmement similaires qui rend vraisemblable le fait que nous possédions tous les mêmes valeurs, ou la même attitude pratique. Même dans ce cas, un tel présupposé est contestable, mais semble relativement vraisemblable. Mais la position de constructiviste universaliste de Korsgaard implique beaucoup plus. Elle implique également de défendre l'idée que tout être rationnel envisageable serait assujetti aux mêmes processus évaluatifs, qu'il s'agisse d'un extraterrestre ou de tout être imaginable. Une telle présupposition semble extrêmement improbable.

A l'inverse de l'attitude extrêmement éthérée défendue par Korsgaard, je pense qu'il est préférable de défendre l'idée que nos attitudes évaluatives quotidiennes passent par des processus beaucoup plus simples. L'idée défendue par Korsgaard que l'évaluation repose toujours sur une forme de valorisation de l'humanité en moi semble en rupture avec l'évidence de nos intuitions morales. Nous ne prenons jamais une telle attitude de rupture totale et d'appel à une identité fondamentale. D'un point de vue phénoménologique, il semble clair que nos valeurs nous semblent justifiées par le fait qu'elles apparaissent comme possédant une certaine justesse et force motivationnelle. Nous pouvons adopter un point de vue réflexif, mais cette réflexion se fait toujours à partir de la mise en avant d'autres valeurs, d'autres choses concrètes auxquelles nous tenons. Une telle perspective est défendue par les tenants du constructivisme humien, et c'est cette perspective alternative que je vais maintenant présenter.

¹¹¹ Korsgaard a bien sûr prévu des objections de ce type, et tente d'y répondre dans la quatrième conférence de *The Sources of Normativity*. Je me contenterai cependant de remarquer que des objections établissent une très faible plausibilité de la théorie de Korsgaard à première vue, et je suppose que le lecteur reconnaîtra ce manque de vraisemblance.

Le constructivisme humien et la justification d'une philosophie expérimentale

Pour comprendre le constructivisme humien, il importe de le distinguer de trois perspectives rivales : le réalisme, le constructivisme kantien à la Korsgaard, et le nihilisme moral. Contrairement au réalisme méta-éthique, le constructivisme humien défend qu'il n'existe pas de valeur existant indépendamment des individus porteurs de valeur. Contrairement au nihilisme, qui nie l'existence de bonnes réponses en morale et qui nie toute possibilité d'objectivité, le constructivisme humien soutient qu'il est possible de trouver de bonnes réponses en morale. Ce double refus constitue le cœur des positions constructivistes en méta-éthique : celles-ci affirment l'existence de réponses correctes d'un point de vue moral, tout en refusant l'idée que ces réponses correctes dépendent d'objets moraux indépendants. Les deux types de constructivisme divergent quant au fait de savoir dans quelle mesure il est possible de définir ce qu'est la « bonne » réponse en matière de morale de façon universelle, pour tous les êtres rationnels, ou dans quelle mesure cette « bonne » réponse doit prendre en compte le point de vue particulier et les idées d'individus particuliers. Contrairement au constructivisme kantien à la Korsgaard, enfin, le constructivisme humien défend qu'il faut considérer les attitudes et les valeurs portées par des individus pour savoir quelle est la bonne action rationnelle dans un certain contexte, sans qu'il soit possible de définir une action normative de façon universelle.

Voilà comment Sharon Street résume la différence entre constructivisme humien et constructivisme kantien : « Les versions humiennes du constructivisme méta-éthique, au contraire, nient que des conclusions substantielles d'un point de vue moral peuvent être déduites du point de vue évaluatif en tant que tel. Selon le humien, le contenu substantiel des raisons normatives d'un agent dépend de son point de départ évaluatif particulier et contingent. Selon ce point de vue, des « raisons pures pratiques », comprises comme le point de vue évaluatif, ou propres au jugement normatif en tant que tel, n'obligent personne à adopter des valeurs substantielles particulières. [...] Le constructiviste humien accepte l'idée que si quelqu'un était entré dans le monde avec un ensemble de valeurs radicalement différentes, [...] alors les raisons normatives de cette personne auraient été, ou deviendraient, radicalement différentes d'une façon correspondante¹¹². »

¹¹² S. STREET, « Coming to Terms with Contingency », *op. cit.*, p. 41.

Dans le cas particulier du constructivisme humien, Sharon Street défend l'idée que les valeurs portées par les individus sont dépendantes de processus évolutifs de sélection naturelle. Dans ce cadre, si les processus évolutifs ont amené différentes espèces à adopter différents types de valeurs, Sharon Street considère qu'elles pourraient avoir des raisons normatives différentes, et qu'adopter le point de vue de l'une ou de l'autre espèce pourrait amener à d'autres conclusions normatives. De même, d'un point de vue plus limité, si deux êtres humains, du fait de leur patrimoine génétique, de leur éducation ou de leur culture, possèdent des valeurs différentes, il n'existe aucune garantie *a priori* qu'ils convergeront vers les mêmes vérités morales.

Il est important de détailler en quoi le point de vue constructiviste de Sharon Street ne se réduit pas à l'idée que tous les désirs, ou tous les points de vue individuels, se valent. Une telle attitude, qui se ramènerait à l'impossibilité de toute critique, condamnerait à l'impossibilité de former une théorie morale normative. Le constructivisme humien défend l'idée qu'il existe bien une vérité en morale, mais que celle-ci est relative à un certain point de vue, et que cette vérité n'est pas réductible aux désirs ou aux attitudes passagères des individus.

Pour savoir ce qu'un individu doit faire à un moment donné, ou ce que serait l'attitude la plus conforme à ses valeurs, il faut prendre en compte la capacité qu'ont les êtres rationnels d'harmoniser leurs différents désirs et de mettre en forme leurs différentes attitudes. Sharon Street donne trois caractéristiques qui permettent de distinguer le fait de valoriser quelque chose du fait de simplement le désirer.

Tout d'abord, selon elle, « l'attitude évaluative est caractérisée par une « discipline » que ne possède pas le simple désir¹¹³. » Cette discipline, selon Street, se caractérise à son tour par deux traits : elle implique une rationalité dans le choix entre fins et moyens, d'une part, et une cohérence entre les différents processus évaluatifs, d'autre part. En d'autres termes, Sharon Street considère qu'accorder de la valeur à quelque chose possède une certaine caractéristique holiste et rationnelle. Holiste, d'une part, parce que la valorisation implique une mise en cohérence entre nos différentes valeurs, ce qui implique que valoriser quelque chose doit être examiné en lien avec ce que nous valorisons par ailleurs. Rationnel, d'autre part, parce que toute valeur a des répercussions sur les moyens utiles pour atteindre nos fins. Si je valorise la cuisine, je dois également accorder un minimum de valeur au temps ou aux capacités nécessaires pour être un bon cuisinier. Contrairement à cette forme d'holisme et de rationalité propre au fait de valoriser quelque chose, les simples désirs ne manifestent pas nécessairement

¹¹³ *Ibid.*, p. 43.

un tel degré de cohérence. Street prend le cas d'un grand blessé, pour lequel l'amputation de la jambe est le seul moyen de garder la vie. Il n'y a rien de contradictoire à dire que ce blessé ne désire en rien être amputé de la jambe, même s'il serait plus difficile de dire qu'il n'accorde aucune valeur au fait d'être amputé de la jambe, alors même qu'il s'agit de la seule solution pour garder la vie. De la même manière, il n'y a aucune difficulté à l'idée d'avoir deux désirs absolument opposés, ou à avoir un désir qui n'est aucunement en cohérence avec les autres désirs d'un individu.

La deuxième caractéristique qui sépare les processus de valorisation des simples désirs, selon Street, consiste dans une phénoménologie plus riche propre aux actes de valorisation. Valoriser quelque chose peut passer par plusieurs types d'émotions différentes. On peut reconnaître que quelqu'un accorde de la valeur à quelque chose par différents signes : par le fait qu'il regrette que cette chose ne se soit pas produite, par le fait qu'il considère que cette chose est importante, etc. Par exemple, quelqu'un peut suivre ce qu'il considère être son devoir en faisant son travail, en s'engageant pour l'armée, ou autre, sans que cette personne ait forcément un désir manifeste d'effectuer ces différentes actions.

La troisième caractéristique qui distingue la valorisation du désir tient à la plus grande complexité de la structure de la valorisation. En effet, la valorisation est prise dans un réseau d'autres valorisations, qui peuvent se renforcer ou s'infirmier mutuellement, alors que le désir peut être parfaitement simple. On retrouve le caractère holiste de la valorisation, opposé à l'atomisme du désir.

A partir de cette triple caractérisation de la valeur, on voit bien en quoi le constructivisme humien ne conduit pas à l'impuissance de la philosophie. En effet, étant donné que les valeurs doivent être soumises à un certain degré de cohérence et de rationalisation, il doit être possible de dire dans quelle mesure un individu se trompe sur ses propres valeurs. Par exemple, si quelqu'un considère que la vie animale n'a aucune valeur, mais en même temps considère que la souffrance est un mal, quelle que soit la créature qui la ressent, alors il est possible de souligner la contradiction entre ces deux énoncés et de proposer une théorie morale qui valorise la vie animale. On peut alors considérer que le rôle de la philosophie est de mettre en ordre les valeurs partagées, afin de produire une nouvelle philosophie publique qui possède une plus grande cohérence que les croyances ordinaires des individus.

Selon la théorie méta-éthique proposée par le constructivisme humien, la philosophie n'est donc pas vaine. Elle peut montrer aux différents individus qu'ils ont tort dans leurs processus de valorisation, et qu'ils feraient mieux d'adopter de nouvelles valeurs, qui sont davantage en cohérence avec leurs désirs ou leurs valeurs fondamentales. Cependant, cette pertinence de la

philosophie ne peut se faire qu'à certaines conditions. Étant donné que la normativité doit forcément partir des attitudes évaluatives portées par les individus, il est absolument nécessaire de savoir quelles sont les choses auxquelles tiennent les individus. Le philosophe ne peut pas philosopher seul, mais doit connaître les attitudes et les choses valorisées par ses concitoyens. Enfin, si le philosophe se trouve en désaccord radical avec la plupart des gens, il n'est en rien évident qu'il ait raison ; il est possible que sa philosophie n'ait pas respecté les valeurs profondes des individus, et qu'il ait échoué à trouver la plus grande cohérence possible compatible avec les idées populaires.

Adopter le constructivisme humien comporte un autre risque. Accepter cette position implique également d'envisager la possibilité que certains individus possèdent des valeurs radicalement différentes, dont il serait alors impossible de démontrer l'irrationalité absolue. Par exemple, il est théoriquement possible que certains individus adoptent une perspective radicalement égoïste, qui n'accorde aucune valeur au bien-être ou aux intérêts d'autres individus. L'existence de personnalités psychopathes correspond à un cas empirique bien connu¹¹⁴. Le constructivisme humien impliquerait dans ce cas de reconnaître que cette indifférence au bien-être général n'est en rien irrationnelle ; il est possible de dire qu'une telle attitude est mauvaise, selon le cadre de valorisation adopté par la plupart des gens, mais il serait difficile de trouver des raisons qui soient susceptibles de prouver aux psychopathes qu'ils devraient adopter la moralité et la valorisation commune¹¹⁵. D'autre part, le constructivisme humien reconnaît la possibilité d'un certain relativisme culturel, même si l'existence d'une histoire évolutionnaire commune implique probablement que la plupart des cultures doivent partager un fondement d'intuitions morales communes¹¹⁶.

La possibilité d'une incompréhension radicale entre personnes ou groupes possédant des valeurs communes ne devrait cependant jamais être prise pour acquise. Dans une démarche à la fois empirique et normative, le constructivisme humien nous invite à une enquête empirique, visant à déterminer les idées défendues par la population, et à construire à partir de cela des recommandations pour une plus grande mise en cohérence des idées morales.

¹¹⁴ Pour une bonne introduction à la question de la psychopathie, voir l'ouvrage classique de Robert D. HARE, *Without conscience: the disturbing world of the psychopaths among us*, New York, Guilford Press, 1999.

¹¹⁵ Dans un autre cadre éthique, c'est également la position que Sidgwick avait été contraint d'adopter dans *The Methods of Ethics*, *op. cit.*

¹¹⁶ Pour une défense de cette position, voir par exemple Nicolas BAUMARD, *Comment nous sommes devenus moraux: une histoire naturelle du bien et du mal*, Paris, Jacob, 2010.

La philosophie comme mise en ordre des intuitions partagées : perspective d'évolution culturelle

Cette mise en cohérence des idées morales par le philosophe a été défendue jusqu'ici de façon statique : un philosophe propose de réviser certaines conceptions, parce qu'elles semblent en contradiction avec d'autres idées qui semblent plus fondamentales. Cependant, on peut considérer que cette pratique de proposition de révision des croyances par le philosophe n'est qu'une reproduction à l'échelle réduite d'amélioration et de révision progressive des croyances à l'œuvre dans toute société. Par exemple, on peut considérer que les réflexions précédentes conduisent à l'acceptation d'une conception dialogique et évolutionnaire de la morale, telle qu'elle a été proposée par des auteurs comme John Dewey, Charles Taylor, ou plus récemment par Philip Kitcher¹¹⁷. Selon les conceptions de Philip Kitcher, la morale doit être comprise comme faisant l'objet d'un processus social de sélection, où les sociétés apprennent progressivement ce qui doit être rejeté au nom de valeurs morales plus fondamentales. On peut penser par exemple à la rejection progressive de l'esclavage, ou au rejet d'idées racistes ou sexistes dans les sociétés modernes. En proposant la révision des idées morales, le philosophe se place donc dans un processus plus large d'évolution et d'amélioration des idées morales¹¹⁸.

Dans cette mission philosophique de mise en ordre des idées morales, le lecteur aura sans doute reconnu la célèbre théorie de l'équilibre réflexif proposée par John Rawls. Dans la *Théorie de la justice*, et ensuite de façon plus précise dans son article de 1974 « The Independence of Moral Theory », John Rawls a défendu que la philosophie morale doit procéder à une mise en cohérence des idées morales ordinaires, qui doit se faire en deux temps. Dans un premier temps, qui correspond à ce que Rawls appelle l'équilibre réflexif étroit, le philosophe doit faire œuvre de psychologue ou de sociologue, et doit chercher à mettre en cohérence les théories morales effectivement présentes dans sa société¹¹⁹. Le philosophe doit

¹¹⁷ John DEWEY, *The collected works of John Dewey. 1932: Ethics*, Carbondale, Southern Illinois Univ. Press, 2008 ; Charles TAYLOR, *The ethics of authenticity*, Harvard University Press, 2018 ; Philip KITCHER, *The ethical project*, Cambridge, Mass., Harvard Univ. Press, 2014.

¹¹⁸ Dans ce cadre, on peut voir le problème associé aux théories contemporaines qui reprochent à la philosophie morale contemporaine son autoritarisme, et sa volonté de trancher de façon anti-démocratique tous les débats moraux importants. Les théories morales contemporaines ne nient pas le débat démocratique, et ne proposent pas une solution qui doit être acceptée une fois pour toutes. Elles prennent simplement part dans le processus évolutionnaire de révision et de transformation des croyances propre à toute société.

¹¹⁹ L'insistance de Rawls sur la ressemblance entre philosophie morale et psychologie est sur ce point tout à fait frappante, en ce que les arguments présentés semblent en grande partie défendre la pratique d'une philosophie expérimentale. Voici par exemple comment Rawls conclut son article : « Nous ne devons pas nous détourner de cette tâche parce qu'elle semble en grande partie appartenir à la psychologie et à la théorie sociale, et non pas à la philosophie. » J. RAWLS, « The Independence of Moral Theory », *op. cit.*

alors effectuer un type d'enquête empirique visant à découvrir les grandes théories morales existantes, et sous-jacentes aux idées morales ordinaires. Rawls ne considère pas qu'il s'agisse d'un simple catalogue d'idées populaires : il s'agit, selon lui, de déjà effectuer une sorte de mise en cohérence des idées morales courantes, afin de garder uniquement des théories populaires plausibles. Cette mise en cohérence passe par la tentative de mettre en adéquation les intuitions morales fondamentales et les théories globales : des intuitions sont examinées, qui permettent de formuler des théories capables d'en rendre compte, théories qui nous amènent ensuite potentiellement à réviser certaines intuitions sous-jacentes (par exemple si nous nous rendons compte que la préférence nationale est incompatible avec le respect dû à tous les êtres humains). Cette démarche d'aller et retour entre les intuitions et les théories globales correspond donc à la recherche d'un équilibre entre différents niveaux du jugement moral. Même si Rawls défend l'idée que le philosophe doit rechercher avant tout une structure morale cohérente pour chacune de ces théories, l'équilibre réflexif étroit, ainsi défini, passe avant tout par une enquête empirique visant à déterminer les intuitions morales populaires.

Une fois que cette recherche a été effectuée, une deuxième phase doit se présenter, que Rawls désigne comme équilibre réflexif étendu. Contrairement à l'équilibre réflexif étroit, l'équilibre étendu est un procédé purement normatif ; il se fonde cependant sur les résultats empiriques obtenus lors de la première étape. Au cours de l'équilibre réflexif étendu, le philosophe examine les différentes théories morales populaires qui ont été mises à jour lors de l'équilibre réflexif étroit. Il les compare ensuite dans un examen rationnel, et examine dans quelle mesure une ou plusieurs des théories peuvent être considérées comme supérieures aux autres. Selon les mots de Rawls, le philosophe « cherche la conception, ou la pluralité de conceptions, qui survivraient à l'examen rationnel de toutes les conceptions plausibles et de tous les arguments raisonnables en leur faveur¹²⁰. » Rawls admet qu'il est possible que plusieurs conceptions survivent, et se trouvent révélées comme dénuées de contradictions. La tâche du philosophe est cependant de rechercher la théorie parvenant à créer l'équilibre le plus cohérent entre nos intuitions fondamentales et nos théories de haut niveau.

La description de l'équilibre réflexif et la tâche que Rawls assigne à la philosophie morale sont donc en parfaite harmonie avec les objectifs de la philosophie morale que nous avons dégagés du constructivisme humien de Sharon Street. Pourquoi avons-nous donc présenté la théorie de Street avant celle, évidemment beaucoup plus connue, de Rawls ? C'est que Rawls refuse, dans « The Independence of Moral Theory » comme dans ses textes ultérieurs, de

¹²⁰ *Ibid.*

prendre parti sur les questions méta-éthiques. Il considère que la théorie morale est indépendante des théories épistémologiques et métaphysiques, en tant que la philosophie morale peut rester neutre vis-à-vis de la question de savoir dans quelle mesure il existe des énoncés moraux vrais. Ainsi, selon Rawls, « L'indépendance que la théorie morale possède vis-à-vis de l'épistémologie découle du fait que la procédure de l'équilibre réflexif ne présuppose pas qu'il existe une conception morale correcte. Il s'agit, si vous voulez, d'un type de psychologie qui ne présuppose pas l'existence de vérités morales objectives¹²¹. »

Il me semble difficile de défendre ce type de position en l'absence d'une théorie méta-éthique qui établirait l'absence de vérités morales objectives. En effet, Rawls se fonde sur un seul argument pour défendre l'indépendance de la théorie morale : il défend l'idée que les théories méta-éthiques sont trop controversées, et qu'il est vain de chercher à attendre qu'elles soient résolues pour commencer à faire de la philosophie morale. Ce pragmatisme présente un certain attrait. Cependant, au lieu de procurer à la philosophie morale une indépendance vis-à-vis des théories méta-éthiques, une telle position place la réflexion normative dans un statut précaire ; s'il était avéré qu'il existe des réalités morales indépendantes de l'esprit, alors tout philosophe serait obligé de considérer que l'équilibre réflexif tel qu'il est préconisé par Rawls serait une procédure vaine. En effet, s'il existait effectivement des vérités morales existantes de façon indépendante, alors il serait préférable de les étudier directement, sans passer par la tâche de l'analyse des conceptions morales ordinaires. En mettant en avant les difficultés rencontrées par les théories réalistes en méta-éthique, et en défendant la position d'un constructivisme humien, nous pensons avoir fourni des arguments visant à fonder une théorie morale normative sur des bases saines. Ces fondements, basés sur l'idée qu'il n'existe pas de réalité morale indépendante, permet de justifier l'exercice de l'équilibre réflexif comme méthode philosophique fondamentale en éthique normative. Et l'équilibre réflexif, tel que nous l'entendons et tel que Rawls l'entendait, est fondamentalement indissociable d'une enquête empirique, à la fois psychologique et sociologique. L'importance philosophique d'une enquête expérimentale est ainsi fondée.

¹²¹ *Ibid.*

Pourquoi les philosophes eux-mêmes devraient faire de la philosophie empirique ?

Avant de clore ce chapitre, je souhaiterais brièvement mentionner une autre question importante. Au-delà de la question de savoir pourquoi des données empiriques devraient avoir une importance pour des théories philosophiques normatives, on peut en effet poser une deuxième question : pourquoi est-ce que des philosophes devraient se préoccuper de collecter eux-mêmes des données empiriques ? Les scientifiques ne sont-ils pas les mieux placés pour effectuer cette tâche ? Je ne tenterai pas ici de défendre l'idée que des philosophes doivent nécessairement effectuer des recherches empiriques. Je considère, cependant, qu'une telle attitude peut être extrêmement fertile, aussi bien pour les sciences sociales que pour la philosophie. Si l'on considère l'apport que les philosophes peuvent faire aux sciences sociales, il est important de considérer le fait que les philosophes sont souvent intéressés par des phénomènes différents que les sociologues ou les psychologues, et sont donc susceptibles d'apporter de nouveaux sujets d'études, et de nouvelles questions, aux disciplines empiriques. Dans la tradition française, on peut par exemple penser aux travaux d'histoire effectués par Foucault, Rancière, Lefort ou Rosanvallon, qui ont considérablement contribué à renouveler l'histoire des idées politiques, ou l'histoire de phénomènes sociaux jusque-là relativement ignorés. Dans le monde anglo-saxon, les philosophes expérimentaux se sont attaqués à de nouveaux domaines jusque-là peu étudiés par les psychologues. Ils ont, par exemple, contribué au renouveau de la psychologie morale, et à l'étude de questions marginales comme l'étude des croyances dans le libre-arbitre ou relatives à la théorie de la connaissance.

Réciproquement, quel peut être l'apport à la philosophie du fait que les philosophes eux-mêmes se préoccupent de questions empiriques ? Si j'ai convaincu le lecteur de l'importance d'obtenir des données empiriques relatives aux croyances normatives des individus, alors il faut bien admettre que des enquêtes empiriques effectuées par des philosophes s'imposent pour deux raisons. Tout d'abord, comme le souligne Rawls¹²², les psychologues et les philosophes ne sont pas forcément préoccupés par les mêmes questions : le problème du libre-arbitre, fondamental pour la philosophie expérimentale, a été relativement peu exploré par la psychologie traditionnelle. Dans ce cadre, le fait que des philosophes effectuent eux-mêmes des enquêtes empiriques est la garantie que des questions d'importance fondamentale pour la philosophie ne

¹²² *Ibid.*

resteront pas inexploitées. D'autre part, si les philosophes doivent s'intéresser aux questions empiriques, il me semble important, d'un point de vue méthodologique, qu'ils se familiarisent eux-mêmes avec la méthode scientifique. Il est en effet extrêmement difficile de juger de la qualité d'une enquête ou d'une expérience sans initiation à la collecte de données, et sans connaissance minimale de théorie statistique. Sans cela, le philosophe risque de se trouver à la merci de travaux de piètre qualité, qui risqueraient de l'induire en erreur. On peut prendre pour exemple la vague de travaux dans les années 2000 dans la tradition anglo-saxonne, qui se sont basés sur la théorie de l'amorçage social (*social priming*) pour discréditer la théorie de la vertu¹²³. Or, la plupart des expériences d'amorçage social ont été discréditées au cours de ces dernières années, remettant en cause les théories philosophiques qui s'étaient reposées sur une base aussi friable. Depuis 2011, la mise en avant de la crise de la reproductibilité en sciences sociales a conduit à une remise en question d'une grande partie des résultats expérimentaux classiques, aussi bien en psychologie qu'en économie ou dans d'autres disciplines. Je pense qu'il est évident qu'un philosophe qui effectue lui-même des expériences ou des enquêtes sera mieux à même de naviguer dans ces débats techniques qu'un philosophe qui se limiterait aux expériences de pensée.

Deux choses sont cependant à noter : premièrement, je ne défends évidemment pas l'idée que la psychologie serait la discipline empirique reine, et qu'il faudrait que l'ensemble des philosophes se tourne vers ce domaine. La tradition française, depuis Tocqueville jusqu'à Rosanvallon, a naturellement montré à quel point une ouverture de la philosophie à la recherche en histoire ou en sciences sociales pouvait se montrer fertile. Deuxièmement, je ne cherche évidemment nullement à remettre en cause l'idée que des philosophes peuvent effectuer des contributions majeures sans effectuer eux-mêmes de recherches empiriques. Le renouveau extraordinaire de la philosophie politique anglo-saxonne dans les années 70 s'est fait à moitié via des philosophes qui effectuaient des recherches empiriques (on peut penser aux recherches historiques de Michael Walzer ou de Charles Taylor, par exemple, ou aux travaux économiques d'un Amartya Sen), et à moitié via des philosophes qui, bien qu'ouverts aux connaissances des sciences sociales, n'effectuaient pas eux-mêmes des études scientifiques. Je me contenterai donc ici d'invoquer un pluralisme dans les méthodes employées par les philosophes.

¹²³ On peut voir par exemple certains travaux de John Doris sur le situationnisme, travaux qui négligent complètement le renouveau de la psychologie de la personnalité qui a eu lieu à partir des années 1980. John M. DORIS, *Lack of character: personality and moral behavior*, Cambridge, Cambridge Univ. Press, 2008.

Conclusion : de la philosophie morale comme activité empirique et dialogique

Ce chapitre nous a permis d'évaluer l'importance des données empiriques pour la constitution d'une philosophie politique normative. Nous avons tout d'abord effectué un bref aperçu des différentes théories méta-éthiques, ce qui nous a permis de mettre en avant qu'une théorie morale plausible devait se fonder sur un constructivisme humien, impliquant de nier l'existence de valeurs morales en dehors de l'esprit, et niant la nécessité d'une convergence universelle autour des mêmes valeurs. La défense d'une telle théorie nous a permis de mettre en avant la tâche de la philosophie comme systématisation et rationalisation des idées morales ordinaires. Dans ce cadre, nous avons repris à notre compte la description que John Rawls a fait de l'équilibre réflexif, indiquant que le philosophe doit procéder à une mise en équilibre des intuitions morales courantes et de théories de hauts niveaux.

Bien évidemment, ce chapitre a traité très rapidement des questions extrêmement complexes. Celles-ci devaient cependant être traitées, même de façon brève, et je pense que la plupart des lecteurs seront d'accord pour dire qu'une perspective réaliste en méta-éthique, ou une perspective kantienne, sont des théories fort peu parcimonieuses qui partent avec un désavantage¹²⁴.

Ce chapitre a cependant laissé de côté un débat crucial pour l'utilité d'une philosophie expérimentale. J'ai jusqu'à présent traité de façon indifférenciée toutes les intuitions possibles, comme s'il n'existait aucun expert en matière d'intuitions philosophiques ou morales. On pourrait cependant objecter qu'il existe un groupe de spécialistes en la matière, et que les seules intuitions qui comptent sont celles produites par ce groupe, et non pas par la plupart des individus. Le chapitre suivant constituera donc un examen de l'expertise des philosophes en matière d'intuitions morales.

¹²⁴ Ce désavantage n'est bien sûr pas insurmontable, et je m'excuse par avance vis-à-vis des lecteurs qui demeureront sceptiques vis-à-vis de mes conclusions.

Appendice : Pourquoi l'évolution a aidé à constituer les idées morales chez l'homme

Dans cette annexe, je présente une brève introduction aux débats en psychologie évolutionniste, afin de démontrer pourquoi cette approche fait sens et est, en réalité, indispensable pour expliquer un certain nombre de comportements humains, notamment dans le domaine de la moralité. Je vais brièvement indiquer quelques arguments en faveur d'une telle perspective. De manière générale, l'idée que des facteurs génétiques et des facteurs évolutifs jouent un rôle important pour expliquer à la fois les capacités humaines et le comportement humain sera mobilisée à plusieurs reprises au cours de cette thèse. Il est donc important que le lecteur soit un minimum familiarisé avec les indices scientifiques du rôle que l'évolution a pu jouer dans la formation des compétences morales humaines. Une telle analyse a été mobilisée dans la présentation des thèmes abordés à la fin du chapitre 3, et sera de nouveau utilisée dans la discussion des débats tournant autour de l'influence génétique sur les talents individuels à partir du chapitre 9. Cela me permettra aussi de mettre en avant pourquoi je ne mobiliserai que de façon limitée des analyses évolutionnistes dans le cadre des analyses de mérite.

Pour que l'évolution soit utile pour expliquer le développement des idées morales, plusieurs présupposés sont nécessaires. Il faut 1) que la génétique ait en général une influence sur le comportement, humain ou animal, 2) qu'il puisse y avoir une pression évolutionniste en faveur du développement de certains types de comportement, et 3) que cette pression évolutionniste se soit exercée en particulier dans le cadre du développement moral.

Pour que l'évolution puisse avoir un impact sur le comportement humain, un présupposé évident est que le comportement doit pouvoir avoir une base génétique. Si les gènes n'avaient aucune influence sur les actions ou le caractère, alors l'évolution n'aurait aucune base pour agir. Dans le cas de l'influence génétique sur le comportement, il est possible de citer de nombreux exemples classiques en éthologie¹²⁵. Par exemple, en 1964, Walter C. Rothenbuhler a découvert qu'un seul gène régula le comportement de certaines abeilles ouvrières, qui avaient tendance à ouvrir des alvéoles si jamais ce gène était présent, et n'ouvraient jamais les alvéoles en l'absence de ce gène¹²⁶. De manière plus générale, le comportement des animaux est extrêmement différent d'une espèce à l'autre : certaines espèces, comme les bonobos, sont

¹²⁵ Pour un très bon résumé des arguments en faveur d'une telle idée, voir le livre de Raghavendra GADAGKAR, *Survival strategies: cooperation and conflict in animal societies*, Harvard University Press, 2001.

¹²⁶ Cité dans le livre de Gadagkar, *Ibid.*

relativement pacifiques, d'autres, comme les chimpanzés, sont beaucoup plus agressifs, et des cas d'extermination complète de certaines bandes sont connus. Expliquer ces différences majeures de comportement sans faire appel à un impact génétique sur le comportement serait hasardeux.

A partir du moment où l'existence d'une influence génétique sur le comportement chez l'animal est démontré, on peut considérer qu'il est *ipso facto* également présent chez l'homme. Le nier serait en effet considérer que l'homme est un empire dans la nature, et serait soustrait aux règles causales générales impliquées dans l'ensemble du monde animal. Cependant, d'autres preuves, plus directes, existent également dans le cas humain. Les indices les plus clairs reposent sur l'existence d'études sur les jumeaux¹²⁷ (les *twin studies*). Il existe deux types principaux d'études sur les jumeaux : les études basées sur la comparaison entre jumeaux élevés dans la même famille, et les études sur des jumeaux élevés dans des familles différentes. Dans le cas des jumeaux élevés dans la même famille, la comparaison se fait entre des vrais jumeaux (donc partageant l'intégralité de leur patrimoine génétique) et des faux jumeaux (qui ne possèdent pas plus de similarité génétique que deux frères et sœurs ordinaires). Dans ce cadre, tout surcroît de similarité entre le comportement des vrais jumeaux, par rapport à la similarité de comportement existant entre des faux jumeaux, peut être attribué à leur plus grande similarité génétique (puisque l'on peut supposer que vrais et faux jumeaux partagent le même milieu familial et social). Dans le cas des jumeaux élevés à distance et ne se connaissant pas, toute similarité entre eux peut être attribuée à des facteurs génétiques. Ces études ont clairement montré un rôle très important des facteurs génétiques dans le développement du comportement et des capacités humaines.

Bien sûr, ces études ne sont pas sans critique possible. Dans le cas de la comparaison entre jumeaux élevés au sein de la même famille, il est possible à la fois que leurs parents ou leur environnement les traite différemment, ou qu'ils s'imitent l'un l'autre, renforçant ainsi leur similarité. Une telle critique est cependant plus délicate dans le cas de jumeaux qui n'ont pas grandi ensemble. Même si des critiques sont toujours possibles, il semble cependant douteux qu'elles parviennent à réduire à néant l'intégralité de l'effet attribué à des facteurs génétiques, tant celui-ci est important.

Le comportement peut donc avoir une base génétique, aussi bien chez l'humain que chez l'animal. Il existe, de plus, de très forts arguments en faveur du fait que l'évolution a eu une

¹²⁷ Pour une introduction à ces questions, voir Valerie S. KNOPIK, Jenae M. NEIDERHISER, John C. DEFRIES et Robert PLOMIN, *Behavioral genetics*, Seventh edition., New York, Worth Publishers, 2017 ; Robert PLOMIN, *Blueprint: how DNA makes us who we are*, Cambridge, Massachusetts, The MIT Press, 2018.

forte influence dans le développement de certains traits moraux. Je ne m'aventurerai pas dans la question de savoir dans quelle mesure l'intégralité des pratiques et des jugements relatifs à la moralité chez l'homme peuvent être directement expliqués par une perspective évolutionniste. Il suffit, en effet, pour défendre les arguments développés par Street, de montrer que certains comportements moraux particulièrement importants chez l'homme possèdent un fondement évolutionniste clair. Nous avons déjà cité dans le corps du chapitre les exemples que prenait Sharon Street elle-même : « (1) Le fait que quelque chose promouvrait notre survie est une raison en sa faveur. (2) Le fait que quelque chose servirait les intérêts d'un membre de notre famille est une raison en sa faveur. (3) Nous avons de plus grandes obligations d'aider nos propres enfants que nous n'en avons d'aider de parfaits étrangers. (4) Le fait que quelqu'un m'a bien traité est une raison de bien traiter cette personne en retour. (5) Le fait que quelqu'un est altruiste est une raison de l'admirer, de le louer, et de la récompenser. (6) Le fait que quelqu'un a causé un tort délibéré est une raison de l'éviter ou de chercher à la punir¹²⁸. » Nous allons maintenant détailler l'origine évolutionniste de ces divers sentiments.

Tous ces exemples reprennent des sentiments très largement partagés d'un point de vue inter-culturel ; or, pour la plupart de ces exemples, la psychologie évolutionniste offre une explication très forte de l'apparition de ces traits, et des raisons pour lesquelles il s'agit de traits populaires dans toutes les sociétés humaines. Dans presque tous les cas, il est assez net que des individus ayant adopté ces comportements ont augmenté les chances que les gènes favorisant ces comportements se transmettent aux nouvelles générations¹²⁹. Le premier cas pris par Street, celui de la survie personnelle, n'est certes pas un cas moral. Les deux exemples suivants illustrent l'importance de favoriser les membres de sa famille, y compris si cela implique de les favoriser aux dépens de personnes qui sont moins proches. Le favoritisme familial est un trait universel, et il est très clair qu'il s'agit d'un trait valorisé moralement, même si ses excès (le népotisme) peuvent être critiqués. Dans ce cadre, il est évident qu'un individu ne disposant pas d'un sentiment de l'importance de favoriser sa famille aurait moins tendance à transmettre ses gènes qu'un individu favorisant sa famille. Les raisons de l'évolution de ces traits sont évidentes, et se retrouvent dans l'ensemble des espèces animales.

¹²⁸ S. STREET, « A Darwinian Dilemma for Realist Theories of Value », *op. cit.*, p. 115.

¹²⁹ Cette formulation un peu alambiquée est nécessaire pour traduire l'idée que l'important est la transmission des gènes favorisant l'apparition d'un certain type de comportement, ce qui peut se produire en favorisant la survie d'individus apparentés, et pas seulement des individus descendant directement du porteur du gène.

Le quatrième exemple, qui traduit des normes de réciprocité, a été bien étudié en psychologie évolutionniste depuis les papiers fondateurs de Trivers au début des années 1970¹³⁰. Il est évident que des pratiques de réciprocité peuvent être avantageuses d'un point de vue évolutionniste si elles conduisent à un bénéfice mutuel parmi deux partenaires dans un échange coopératif. Les dernières années ont vu une explosion des études portant sur les pratiques de réciprocité et de mutualisme dans le monde animal, les exemples les plus connus concernant notamment les échanges d'épouillage chez les primates, qui se font effectivement dans le cadre d'un échange réciproque¹³¹.

Le cas des pratiques de mérite évoquées par Sharon Street dans ses exemples 5 et 6 est plus délicat à expliquer d'un point de vue évolutionniste, pour plusieurs raisons. Avant de détailler les deux raisons principales, il est important de noter que la validité d'une analyse en termes de psychologie évolutionniste ne dépend pas du fait que la théorie serait capable d'analyser de façon indiscutable n'importe quel type de phénomène comme une adaptation à un environnement. Comme le reconnaissent la plupart des théoriciens de l'évolution, de nombreux traits dans le monde animal ne sont pas en eux-mêmes des adaptations, mais sont simplement les effets secondaires d'autres adaptations, ou sont de simples traits ayant émergé à la suite de processus stochastiques. Adopter une approche adaptative en biologie ne veut évidemment pas dire que tout doit être considéré comme une adaptation. Je tiens à souligner qu'il existe à la fois des traits pour lesquels une explication en termes évolutionnistes est très facile, comme la préférence pour les parents ou les pratiques de réciprocité ; et des traits qui peuvent être légèrement plus difficiles à expliquer en termes évolutionnistes, comme l'idée de l'importance de récompenser le mérite des personnes tierces.

La première difficulté pour expliquer le désir de récompenser les individus selon leur mérite tient au fait qu'il est difficile de détecter des principes de mérite au sein du monde animal. Même si des éthologues ont pu soutenir que des pratiques similaires existent par exemple chez le chimpanzé ou de nombreux primates, ces expériences ont été fortement critiquées, notamment du fait des difficultés à les reproduire. Dans ce contexte, supposer qu'il existe des pratiques de justice ou de mérite en dehors de l'homme relève encore de la spéculation non fondée sur des preuves solides¹³². Or, en l'absence de comparaison entre différentes espèces, il

¹³⁰ Voir l'article classique de Robert L. TRIVERS, « The Evolution of Reciprocal Altruism », *The Quarterly Review of Biology*, 1971, vol. 46, n° 1, p. 35-57.

¹³¹ Frans B. M. de WAAL, *Peacemaking among primates*, Cambridge, Mass., Harvard Univ. Press, 1989.

¹³² Pour une référence favorable à l'idée d'un sens de la justice chez certains primates, voir notamment Christophe BOESCH et Hedwige BOESCH-ACHERMANN, *The chimpanzees of the Tai Forest: behavioural ecology and evolution*, Oxford, Oxford Univ. Press, 2008. Pour une perspective critique, voir Michael TOMASELLO, *Why we cooperate*, Cambridge, Mass., MIT Press, 2009.

devient beaucoup plus délicat de détecter avec certitude pourquoi ce trait est apparu chez l'homme. Nous ne disposons, de fait, que d'un seul exemple à partir duquel comprendre l'apparition de la moralité. Pour cette raison, les hypothèses évolutionnistes relatives à l'apparition du sens de la justice restent, en grande partie, des spéculations, même s'il s'agit de spéculations qui peuvent être informées par des exercices en modélisation ou par une connaissance (approximative) du mode de vie humain ancestral.

Une deuxième raison complique les explications évolutionnistes des sentiments de mérite, au moins dans le cas des jugements relatifs au fait qu'un individu mérite d'être puni. Cette difficulté repose sur le fait qu'il existe un coût immédiat à punir : si je punis un criminel, je cours le risque que cet individu se venge, et je subis ainsi un dommage potentiel. Pour bien comprendre en quoi ce coût lié à la punition représente une difficulté pour une théorie évolutionniste du mérite, il faut faire la distinction entre punition de la deuxième partie et punition de la troisième partie¹³³ (*second-party punishment* et *third-party punishment* en Anglais). La punition de la deuxième partie concerne des cas où un individu a été attaqué, et riposte ; dans cette situation, la première partie est l'agresseur, la deuxième partie est la personne agressée. On pourrait décrire ceci comme un cas simple de « réciprocité négative », et dans ce cadre, il n'y a aucun problème à expliquer cela dans une perspective évolutionniste. En effet, un individu qui ne se défendrait pas étant attaqué risquerait d'attirer sur lui de nombreuses attaques à l'avenir. Le fait qu'il existe un coût lié à la punition est bien plus que compensé par l'avantage qu'il y a à obtenir une plus grande sécurité à l'avenir.

A l'inverse, la punition de la troisième partie est beaucoup plus délicate à expliquer d'un point de vue évolutionniste. Ce type de punition renvoie à des situations où une troisième personne assiste à un cas d'infraction ou de crime sans en être elle-même la victime, et intervient pour punir le ou la coupable. Il s'agit donc d'un type de punition altruiste ou désintéressé. Or, aucun comportement ne saurait être complètement altruiste d'un point de vue génétique : si un comportement a évolué pour des raisons d'adaptation, c'est nécessairement qu'il fournissait un avantage d'un point de vue de la transmission des gènes, et qu'il ne représentait pas un coût net de ce point de vue.

Dans nos sociétés, la punition de la troisième partie n'est pas vraiment altruiste : elle est déléguée à des spécialistes, participant à la police ou au système judiciaire, qui ont un intérêt professionnel à punir les contrevenants. Dans le cas de ces spécialistes, le risque lié à

¹³³ Je me concentrerai ici sur le cas de la punition, parce que c'est l'exemple le plus étudié dans la littérature en psychologie évolutionniste sur la question ; mais des problèmes similaires se poseraient dans le cas de la récompense.

l'interpellation ou à la punition des criminels est compensé par le salaire qu'ils obtiennent. Cependant, comme l'ont montré aussi bien l'anthropologie que les recherches contemporaines en psychologie culturelle, des intuitions liées à la punition de contrevenants sont très largement partagées au sein de nombreuses cultures, même dans des cas où il n'existe pas de structure étatique séparée capable de faire appliquer la loi. Dans ce cadre, il est beaucoup plus difficile d'expliquer pourquoi les individus ont l'intuition qu'il est important de punir le contrevenant, alors même qu'ils ne sont pas directement victimes du crime. On pourrait penser que de telles intuitions sont contre-productives, en tant qu'elles risquent de mettre en danger l'individu ayant des intuitions de mérite s'il cherche à intervenir dans des disputes où il n'est pas directement impliqué.

De nombreuses théories ont été proposées pour expliquer pourquoi les humains, à travers différentes sociétés, tendent à universaliser leurs jugements de mérite, c'est-à-dire sont amenés à juger que des actions négatives doivent être punies, et des actions positives doivent être récompensées, même dans des cas où ils ne sont pas directement concernés. Par exemple, le psychologue évolutionniste Nicolas Baumard a fait l'hypothèse que nous avons tendance à énoncer des principes de justice universellement valides pour témoigner du fait que nous sommes nous-mêmes de bons coopérateurs, c'est-à-dire que nous n'aurons pas tendance à exploiter ou agresser d'autres individus¹³⁴. D'autres psychologues évolutionnistes, comme Robert Boyd ou Herbert Gintis, ont spéculé que les jugements de mérite peuvent avoir une fonction dans le cadre d'institutions qui réduisent le coût lié à la punition¹³⁵.

Je mentionne ces idées simplement pour témoigner du fait que l'évolution des idées de mérite représente un phénomène extrêmement complexe, et pour expliquer pourquoi je ne m'aventurerai pas sur ce point ici. Cependant, à côté de principes pour lesquels une explication évolutionniste resterait nécessairement complexe et hasardeuse, comme les cas de mérite, on peut trouver des exemples de cas où les jugements des individus s'expliquent simplement par la théorie de l'évolution, comme les cas de réciprocité ou de favoritisme vis-à-vis des membres de sa famille. Dans ce cadre, prendre une perspective évolutionniste reste de toute façon essentiel pour expliquer la structure de la pensée morale humaine.

¹³⁴ N. BAUMARD, *Comment nous sommes devenus moraux*, *op. cit.*

¹³⁵ Robert BOYD, Herbert GINTIS et Samuel BOWLES, « Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare », *Science*, 2010, vol. 328, n° 5978, p. 617-620.

Chapitre 5

La question de l'expertise philosophique

Le chapitre précédent a mis en avant qu'une philosophie normative doit se fonder sur les intuitions morales des individus. Une telle conclusion laisse donc entendre qu'il serait bon de chercher à déterminer quelles sont les conceptions morales du plus grand nombre. Face à cette idée, il est cependant possible de répondre qu'une partie de la population, les philosophes, constituent des experts quant aux intuitions morales. Plus généralement, on peut penser que les individus diffèrent dans leur capacité à connaître ce qui est bon et juste, et qu'il serait illusoire de penser que tous les individus sont capables d'avoir des intuitions pertinentes pour déterminer en quoi consiste la justice ou la moralité.

La question de l'expertise philosophique a constitué le cœur des débats en métaphilosophie au cours des vingt dernières années. De nombreux philosophes, notamment parmi ceux qui ont été sceptiques par rapport au projet d'une philosophie expérimentale, ont adopté une perspective simple : il n'y aurait, selon eux, aucun intérêt philosophique spécifique à obtenir des données empiriques relatives aux croyances populaires quant à la justice, la moralité, ou tout autre concept étudié par les philosophes. Le fait qu'une croyance soit populaire, ou que les individus soient divisés relativement à un sujet philosophique, ne nous apprend absolument rien quant à savoir ce qui constitue la conception philosophique correcte. Les philosophes possèderaient une compétence particulière pour traiter des questions morales ou philosophiques, que ne possèdent pas les gens ordinaires. Cette compétence naitrait d'une réflexion approfondie et prolongée sur un sujet et ses thèmes apparentés. Cette ligne de défense a, sous des formes variées, été défendue par des critiques de la philosophie expérimentale comme Timothy Williamson ou Kirk Ludwig¹³⁶.

¹³⁶ T. WILLIAMSON, « Philosophical Expertise and the Burden of Proof », *op. cit.* ; Kirk LUDWIG, « The Epistemology of Thought Experiments: First Person versus Third Person Approaches », *Midwest Studies In Philosophy*, 2007, vol. 31, n° 1, p. 128-159.

Ce chapitre consistera en un examen général de la possibilité d'une expertise en matière d'intuitions et de théories morales. Étant donné que ces questions ont été centrales dans le développement de la philosophie expérimentale au cours des vingt dernières années, il s'agira aussi d'effectuer une introduction aux débats récents dans ce domaine. Je procéderai en trois temps. Dans un premier temps, je résumerai les débats relatifs aux buts de la philosophie expérimentale. Dans un deuxième temps, je présenterai les résultats de diverses expériences qui tendent notamment à montrer que les philosophes sont aussi affectés par des biais de raisonnement. Enfin, dans un troisième temps, j'effectuerai une défense de la compétence morale de la foule. A défaut de montrer que la vérité morale peut être atteinte par la considération des opinions populaires, je tenterai de montrer qu'un philosophe doit prendre au sérieux les opinions courantes.

Le programme positif et le programme négatif de la philosophie expérimentale

Pour comprendre la diversité des manières de mettre en avant l'utilité des données empiriques, il importe d'abord d'effectuer une typologie des différents buts associés à la philosophie expérimentale. La distinction la plus courante entre les différents programmes liés à la philosophie expérimentale consiste à distinguer un programme positif d'un programme négatif. Je reprendrai ici les distinctions effectuées par Alexander, Weinberg et Mallon dans leur article de référence « Accentuate the Negative¹³⁷ ».

Le programme positif consiste à défendre une position philosophique à partir du fait que cette position semble la position la plus populaire au sein d'une population. A l'opposé, le programme négatif cherche à nier la validité des méthodes philosophiques classiques, faisant appel aux expériences de pensée, en mettant en avant deux facteurs qui, selon les avocats du programme négatif, sont prévalents dans toute réflexion philosophique : la diversité des intuitions en question, et le fait que ces intuitions sont souvent influencées par des facteurs non-pertinents ou par des sources claires de biais.

Ces deux approches ont des objectifs opposés, et le programme négatif remet en cause aussi bien les analyses philosophiques classiques que l'usage de données empiriques pour afficher le

¹³⁷ J. ALEXANDER, R. MALLON et J. M. WEINBERG, « Accentuate the Negative », *op. cit.*

soutien à des positions philosophiques. En effet, si le programme négatif réussit à montrer que les intuitions philosophiques ne sont pas fiables et sont influencées par des facteurs non pertinents, ceci remettrait aussi bien en cause les intuitions des philosophes que les intuitions des non-philosophes, sur lesquelles le programme positif souhaiterait s'appuyer. Le programme négatif défendrait ainsi l'importance d'un scepticisme modéré quant à notre capacité à trancher les débats philosophiques par le raisonnement.

Les auteurs défendant un programme négatif mettent en avant que les individus ordinaires sont souvent divisés dans leurs intuitions, et sont souvent influencés par des facteurs non-pertinents. Cependant, même si l'on admet ces deux faits, ils ne remettraient en cause les jugements effectués par les philosophes que si ceux-ci ne possèdent pas une capacité spéciale qui leur permettrait de fournir de meilleurs jugements que les individus ordinaires. Cette compétence spéciale pourrait se fonder sur deux capacités spécifiques des philosophes : ils pourraient être moins biaisés dans leurs raisonnements que les individus ordinaires, et ils pourraient baser leurs jugements sur une plus grande réflexion. Pour cette raison, les tenants du programme négatif ont cherché à montrer que les philosophes ne possèdent pas de compétence spéciale, sur la base de deux arguments : d'une part, les philosophes sont aussi biaisés que les individus ordinaires, et d'autre part, la réflexion ne conduit que très rarement à un changement d'opinion quant à la solution d'un problème philosophique. Je vais maintenant présenter ces différentes expériences visant à démontrer l'existence de biais de raisonnements chez les philosophes.

Y a-t-il un biais des philosophes ? La susceptibilité aux effets d'ordre

L'idée que les philosophes seraient aussi biaisés que les individus ordinaires possède une assez forte base empirique, si on interprète cette idée comme un cas particulier de l'idée que les experts sont soumis aux mêmes biais et heuristiques psychologiques que les profanes. En effet, de nombreuses études psychologiques classiques ont démontré que des experts, dans des domaines très divers, étaient susceptibles d'être affectés par une multiplicité de biais courants. Par exemple, des études classiques ont montré que les agents immobiliers étaient soumis à l'effet d'ancrage dans leurs estimations de la valeur d'un logement, ou que les juges étaient

soumis à ce biais dans leurs évaluations des peines, de la même manière que des individus naïfs¹³⁸.

Dans le cadre spécifique des philosophes, quelques études ont montré que les philosophes ont tendance à montrer certains biais de la même manière que des non-philosophes. Par exemple, il s'agit d'un constat classique et bien établi que les individus sont affectés par des effets d'ordre, de telle sorte que l'ordre de présentation des expériences change parfois les réactions des individus.

Le philosophe Eric Schwitzgebel et le psychologue Fiery Cushman ont comparé la susceptibilité des philosophes et des non-philosophes aux effets d'ordre au cours d'une série d'expériences. Dans un article publié en 2012¹³⁹, ils ont présenté à divers participants une série de questions morales qui étaient, selon eux, susceptibles de générer des effets d'ordre. Leur échantillon se composait de 324 philosophes (qui avaient obtenu ou bien un master ou bien une thèse en philosophie), 753 universitaires non-philosophes, et 1389 personnes qui n'avaient obtenu ni master ni thèse et qui avaient été recrutées via un site internet. Les scénarios posant des dilemmes philosophiques utilisés par Schwitzgebel et Cushman peuvent se ranger en trois grandes catégories : des scénarios mettant en œuvre la doctrine du double effet (dont l'exemple le plus connu est le dilemme du tramway¹⁴⁰), des scénarios testant la différence de jugement

¹³⁸ Gregory B NORTHCRAFT et Margaret A NEALE, « Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions », *Organizational Behavior and Human Decision Processes*, 1987, vol. 39, n° 1, p. 84-97 ; Birte ENGLISH, Thomas MUSSWEILER et Fritz STRACK, « Playing dice with criminal sentences: the influence of irrelevant anchors on experts' judicial decision making », *Personality & Social Psychology Bulletin*, 2006, vol. 32, n° 2, p. 188-200 ; Daniel KAHNEMAN et Gary KLEIN, « Conditions for intuitive expertise: a failure to disagree », *The American Psychologist*, 2009, vol. 64, n° 6, p. 515-526.

¹³⁹ Eric SCHWITZGEBEL et Fiery CUSHMAN, « Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers », *Mind & Language*, 2012, vol. 27, n° 2, p. 135-153.

¹⁴⁰ Le dilemme du tramway est tellement débattu dans la philosophie contemporaine que je me contenterai de rappeler en bas de page l'essentiel du dilemme. Le dilemme trouve son origine dans deux articles publiés par Philippa Foot et Judith Jarvis Thomson. Philippa FOOT, « The Problem of Abortion and the Doctrine of Double Effect », *Oxford Review*, 1967, vol. 5, p. 5-15 ; Judith Jarvis THOMSON, « Killing, Letting Die, and the Trolley Problem », *The Monist*, 1976, vol. 59, n° 2, p. 204-217. Voici les principales caractéristiques du dilemme. Imaginez que vous vous promenez au bord d'une voie de chemin de fer, quand vous apercevez soudain un tramway sans contrôleur fonçant à toute vitesse. Vous voyez que, si vous ne faites rien, le tramway va écraser cinq personnes qui se trouvent plus loin sur la voie. À côté de vous se trouve cependant un levier, qui vous permettrait de détourner le tramway sur une voie de garage. Si vous actionnez le levier, cependant, une personne se trouvant sur cette voie de garage sera écrasée. Est-il acceptable d'un point de vue moral de détourner le tramway ? À cette question, la majorité des individus interrogés répondent oui. Les individus semblent donc considérer qu'il peut être acceptable de sacrifier un individu pour en sauver cinq. Cependant, une autre version du dilemme donne de tout autres résultats. Imaginons maintenant que vous vous trouvez sur une passerelle au-dessus d'une voie de chemins de fer ; de nouveau, vous voyez un train sans conducteur arriver à toute vitesse sur la voie en dessous de vous. Sans votre action, elle va écraser à coup sûr cinq personnes se trouvant sur les voies. Cependant, à côté de vous se trouve un homme très gros ; vous savez que, si vous le poussez sur la voie, il arrêterait à coup sûr le train, ce qui sauverait les cinq personnes. Est-il moralement acceptable de pousser le très gros homme ? À cette question, la grande majorité des individus répondent non. Expliquer les différences de jugement entre ces deux cas a occupé une grande partie des philosophes et psychologues moraux contemporains.

entre des cas d'action ou d'omission¹⁴¹, et des scénarios testant l'impact de cas de fortune morale sur des jugements de blâme¹⁴². Dans ces trois types de scénarios, de nombreuses expériences ont démontré que les individus jugent, en moyenne, le fait d'actionner le levier dans le dilemme du tramway comme moins mauvais que de pousser une personne du haut du pont, jugent les cas d'omission comme moins mauvais d'un point de vue moral que les cas d'action, et jugent qu'un individu est moins blâmable s'il a été plus chanceux et n'a pas causé d'accident. Dans tous ces cas, les chercheurs voulaient savoir si présenter un type de scénario avant l'autre changerait les jugements moraux des individus, et si les philosophes et non-philosophes seraient susceptibles de la même manière à ces effets d'ordre.

Schwitzgebel et Cushman ont trouvé de clairs effets d'ordre aussi bien dans le cas du dilemme du tramway que dans le cas des scénarios impliquant des cas d'action et d'omission¹⁴³. Par exemple, dans le cas du dilemme du tramway, les individus devaient noter l'action sur une échelle de 1 à 7, 1 indiquant « extrêmement bon d'un point de vue moral », et 7 indiquant « extrêmement mauvais d'un point de vue moral ». Si des individus voient d'abord le dilemme original (où le spectateur doit actionner un levier), et ensuite seulement le dilemme du pont, alors la réponse moyenne est de 3,38 pour le cas du levier, et de 4,13 pour le cas du pont. Par contre, si les individus voient d'abord le dilemme du pont, alors la réponse moyenne des participants dans le cas du levier passe à 3,88, alors que la réponse moyenne pour le cas du pont monte à 4,33. En d'autres termes, les différences sont relativement faibles si l'on ne prend en compte que le deuxième scénario vu par les participants (3,88 pour le cas du levier, 4,13 pour le cas du pont), alors que les différences sont énormes si l'on se concentre dans le cas du premier scénario (3,38 contre 4,33). Des différences fortes sont aussi constatées dans le cas des jugements d'action et d'omission. Or, les philosophes, les universitaires non-philosophes, et les non-universitaires témoignent des mêmes effets d'ordre dans tous les scénarios : il semble que

¹⁴¹ Comme exemple de scénarios testant la différence entre action et omission, on peut prendre la différence entre prendre la bouée de sauvetage de quelqu'un lors d'un naufrage et refuser de donner sa bouée de sauvetage à quelqu'un qui n'en aurait pas.

¹⁴² Un cas de fortune morale implique des scénarios où deux agents ayant effectué une action extrêmement similaire, avec le même état d'esprit, se trouvent causer des états de choses radicalement différents par le pur effet du hasard. L'exemple classique compare deux personnes ayant pris la route dans un état d'ivresse, et ayant exactement les mêmes capacités. Dans un cas, la personne rentre à la maison sans encombre. Dans l'autre, le conducteur est dans l'incapacité d'éviter un enfant traversant la route. On peut avoir l'intuition que le conducteur ayant écrasé l'enfant mérite un blâme bien plus grand, alors même que son action est extrêmement similaire à celle de l'autre personne.

¹⁴³ Les individus réagissaient à l'inverse de façon similaire quel que soit l'ordre des cas de fortune morale.

l'expertise philosophique ne change en rien la réaction à ce type d'effet. Un autre article des mêmes auteurs, publié en 2015, confirme ces résultats, en se servant de nouveaux scénarios¹⁴⁴.

Il est très difficile d'expliquer ce type d'effet d'ordre par un raisonnement rationnel ; en théorie, les réponses d'un individu expert devraient être stables, et ne devraient pas subir une influence forte de la part de cas différents qu'il vient de considérer. Cela montre que des facteurs qui devraient ne pas être pertinents ont tout de même un impact sur les jugements moraux des individus. Ce type de résultats est d'autant plus surprenant que les philosophes ont certainement une familiarité bien plus importante avec les scénarios proposés que les non-philosophes ; contrairement à ces derniers, ils devaient bien être capables d'imaginer les autres cas pertinents. Un philosophe considérant le dilemme du levier ou du pont a certainement à l'esprit l'autre scénario, et est capable de se rendre compte qu'évaluer l'acte de pousser la personne grosse ou l'acte de tirer le levier doit être évalué en comparaison avec d'autres scénarios. Malgré cet avantage, les philosophes ont fait preuve de la même susceptibilité à un facteur non-pertinent, à savoir l'ordre de présentation des scénarios. Il s'agit là d'un indice que l'expertise philosophique ne soustrait pas aux biais ordinaires, même dans la considération de problèmes philosophiques classiques.

Un deuxième argument contre l'expertise des philosophes : le manque d'impact de la réflexion sur les intuitions morales

Le premier argument pour réfuter la compétence spéciale des philosophes se fonde donc sur le fait qu'ils sont susceptibles de tomber dans les mêmes biais de raisonnement que les non-philosophes. Bien sûr, le fait que les philosophes peuvent fournir des jugements biaisés n'est pas en soi suffisant pour nier toute expertise. En conséquence, un deuxième argument a été mis en avant par Edouard Machery et ses collaborateurs pour réfuter que les philosophes disposent d'une compétence spéciale : Machery et ses collègues défendent l'idée qu'un surcroît de réflexion conduit très rarement à un changement d'opinion dans le cadre de l'analyse de

¹⁴⁴ Eric SCHWITZGEBEL et Fiery CUSHMAN, « Philosophers' biased judgments persist despite training, expertise and reflection », *Cognition*, 2015, vol. 141, p. 127-137.

problèmes philosophiques, et qu'il est donc vain de penser que les philosophes seraient plus compétents sur un sujet parce qu'ils y auraient davantage réfléchi¹⁴⁵.

Nier l'importance de la réflexion est évidemment une idée extrêmement forte, qui demande une base empirique solide. La critique de la réflexion par Machery et ses collègues prend place dans le cadre d'une théorie qui s'est considérablement développée au cours de ces vingt dernières années, l'intuitionnisme moral. Cette théorie, défendue notamment par le psychologue américain Jonathan Haidt, défend que les idées des individus relatives aux questions morales sont avant tout le résultat de réactions intuitives, sans que les individus aient accès aux facteurs qui expliquent l'apparition de ces réactions intuitives. Des facteurs a-rationnels influenceraient nos jugements moraux ; la réflexion ne servirait alors qu'à l'élaboration de justifications post-hoc, qui seraient sans rapport avec les vraies causes à la base de nos jugements moraux. Même si l'intuitionnisme moral accepte bien que le raisonnement puisse influencer nos jugements moraux, Haidt considère qu'il s'agit un cas relativement rare en l'absence d'un débat public intense.

Dans une étude influente écrite en 2000, Jon Haidt présente les résultats d'une expérience où il avait demandé à des participants de justifier leur réaction face à un cas d'inceste consensuel entre un frère et une sœur, où toutes les précautions avaient été prises pour qu'aucune conséquence néfaste ne résulte de la relation sexuelle¹⁴⁶. La plupart des individus interrogés ont condamné la relation incestueuse ; cependant, confrontés à la nécessité de justifier leur décision, presque tous les individus ont fait appel aux conséquences négatives possibles de l'acte, alors même que ces conséquences possibles avaient été décrites comme impossibles dans le scénario qu'ils avaient lu. Jon Haidt en a conclu que les individus avaient un jugement intuitif de dégoût par rapport à l'inceste, qui était véritablement à l'origine de leurs jugements moraux, alors même que les justifications des individus faisaient appel à des facteurs qui n'étaient pas pertinents pour le cas en question. A partir de ce cas, il en a conclu à la faible pertinence de la réflexion et de la rationalité pour l'exercice du jugement moral¹⁴⁷. Il va sans dire que, si cette vision du rôle de la rationalité dans l'exercice du jugement moral était vraie, alors l'idée

¹⁴⁵ David COLAÇO, Markus KNEER, Joshua ALEXANDER et Edouard MACHERY, « On second thought: A refutation of the reflection defense », *manuscrit non publié*.

¹⁴⁶ Jonathan HAIDT, Fredrik BJÖRKLUND et Scott MURPHY, « Moral Dumbfounding: When Intuition Finds No Reason », *manuscrit non publié*, p. ; Jonathan HAIDT, « The emotional dog and its rational tail: A social intuitionist approach to moral judgment. », *Psychological Review*, 2001, vol. 108, n° 4, p. 814-834.

¹⁴⁷ Pour une critique puissante de l'expérience de Haidt montrant que les conclusions de l'auteur ne suivent pas des résultats expérimentaux, voir Edward ROYZMAN, Kwanwoo KIM et Robert LEEMAN, « The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect », *Judgment and Decision Making*, 2015, vol. 10, p. 296-313.

d'équilibre réflexif serait sérieusement en péril, et il deviendrait difficile de défendre l'idée que les philosophes pourraient apporter une aide décisive à la pensée morale.

Avant de voir pourquoi une conclusion aussi pessimiste n'est sans doute pas justifiée, voyons les expériences utilisées par Machery et ses collègues pour défendre l'idée que le rôle des intuitions a-rationnelles n'est pas limité au jugement moral, mais s'étend à l'ensemble des cas étudiés par les philosophes. Dans un article écrit en 2016, David Colaço, Markus Kneer, Joshua Alexander et Edouard Machery ont présenté une série de problèmes philosophiques à plus de 2500 participants¹⁴⁸. Les différentes questions philosophiques utilisées renvoyaient à un cas de Gettier¹⁴⁹, au cas de la Virgin Queen¹⁵⁰, à un cas de Gödel¹⁵¹, et à l'effet Knobe¹⁵². Pour chacun des différents scénarios, les auteurs ont utilisé différentes conditions visant à stimuler la réflexion des participants, et ont comparé les réponses à celles de participants dans une condition de contrôle. Quatre types de conditions différentes ont tenté de stimuler la réflexion. Dans un premier cas, les participants ne pouvaient pas donner leur réponse à chacun des scénarios immédiatement, mais étaient contraints d'attendre 40 secondes avant de pouvoir soumettre leur réponse. Dans un deuxième cas, les participants devaient justifier leur réponse, et étaient informés à l'avance qu'ils seraient contraints de fournir une justification. Dans une troisième condition, les participants apprenaient qu'ils pourraient obtenir une récompense s'ils fournissaient la « bonne » réponse. Dans une quatrième condition, les participants devaient d'abord résoudre des problèmes mathématiques afin de stimuler leurs capacités analytiques.

Chacune de ces méthodes peut être critiquée, en ce qu'il n'est pas forcément évident qu'il s'agisse d'un bon moyen de stimuler la réflexion. Cependant, il est hautement probable qu'au moins une d'entre elles joue un rôle efficace pour augmenter la réflexivité des participants. Or, le résultat obtenu par Colaço et ses collègues est net : aucune des manipulations n'a d'effet sur les réponses des individus, qui sont dans tous les cas hautement similaires aux réponses obtenues dans le cas où aucun moyen n'est utilisé pour stimuler la réflexivité des participants.

¹⁴⁸ D. COLAÇO, M. KNEER, J. ALEXANDER et E. MACHERY, « On second thought », *op. cit.*

¹⁴⁹ Le cas de Gettier renvoie à une situation où un individu possède une croyance vraie et justifiée, mais où cette croyance est vraie par accident. Le problème philosophique est alors de savoir dans quelle mesure il s'agit d'un authentique cas de connaissance.

¹⁵⁰ Le cas de la Virgin Queen renvoie à une situation où un individu est capable de dire la bonne réponse à une question (dans ce cas, la date de la mort de la reine Elizabeth), mais doute tellement de sa réponse qu'il a le sentiment que cette réponse n'est pas la bonne. La question est alors de savoir si cet individu peut connaître la réponse sans croire que cette réponse est vraie.

¹⁵¹ Dans le cas de Gödel, il s'agit de savoir dans quelle mesure, si jamais l'auteur du théorème de l'incomplétude de Gödel n'avait pas été Gödel, mais un ami de Gödel disparu dans de mystérieuses circonstances, le terme Gödel renverrait ou bien à l'individu Gödel, ou bien à l'auteur véritable du théorème d'incomplétude.

¹⁵² L'effet Knobe renvoie à la question de savoir dans quelle mesure un patron d'entreprise qui lance un plan économique ayant pour effet secondaire de détériorer l'environnement a intentionnellement détérioré l'environnement.

Les auteurs concluent que la réflexion n'a pas d'impact majeur sur les intuitions liées aux problèmes philosophiques fondamentaux.

Cette expérience est hautement intéressante, et montre qu'un surcroît de réflexion n'a pas toujours d'impact sur les réponses fournies à des problèmes philosophiques. Je pense qu'une telle expérience ne montre pas, par contre, que la réflexion n'a aucun impact sur la qualité des réponses philosophiques apportées.

Trois raisons en faveur de l'importance de la réflexion pour les intuitions philosophiques

La conclusion de Colaço et de ses collègues, selon laquelle la réflexion n'a pas d'impact sur les réponses philosophiques, est extrêmement ambitieuse, et contestable. Je vais examiner trois types d'objections possibles à l'expérience de Colaço et de ses collègues. Deux types de réponses viennent immédiatement à l'esprit : tout d'abord, contester l'efficacité de la réflexion semble relativement aberrant, en ce que la réflexion semble avoir un impact positif évident dans la vie quotidienne. Ensuite, on pourrait également objecter que la réflexion supplémentaire créée par ces conditions expérimentales n'est pas la même que la réflexion produite par une activité philosophique authentique. Colaço et ses collègues ont prévu ces objections, et ont tenté d'y répondre.

D'après Colaço et ses collègues, il n'est pas évident que la réflexion puisse jouer le même rôle dans la vie quotidienne et dans l'examen d'expériences de pensée philosophiques. En effet, lorsque nous nous servons de la réflexion dans la vie quotidienne, c'est la plupart du temps pour examiner des questions où il existe une bonne réponse. Par exemple, si je réfléchis un peu plus longtemps pour résoudre un problème de mathématique, alors j'exerce mes capacités afin d'obtenir de meilleures chances de trouver une réponse correcte, et pour laquelle je suis sûr qu'il existe une bonne réponse. A l'inverse, il n'est pas sûr que, dans le cas des expériences philosophiques citées, les participants considèrent qu'une réponse est nécessairement meilleure qu'une autre, ou qu'il existe une bonne réponse en général. Ainsi, comme le soulignent Colaço et ses collègues, bien qu'il existe une solide littérature en psychologie et en économie comportementale soutenant que la réflexion améliore les performances des sujets, cela ne veut en rien dire que la réflexion devrait améliorer les performances des sujets quand il n'existe pas de bonne réponse évidente. « Les jugements que font les gens dans les types de situations

examinées par les psychologues sociaux et les économistes comportementaux sont erronés à la lumière des conceptions que les gens se font eux-mêmes ; quand cela se produit, les gens peuvent changer leur jugement quand on leur offre la possibilité de réfléchir. Par contraste, quand les jugements sont faits avec confiance et ne sont pas erronés à la lumière des conceptions des participants, la réflexion n'influence pas le jugement ; elle conduit plutôt les gens à penser à des arguments en faveur des jugements qu'ils ont effectués de façon indépendante¹⁵³. »

Passons maintenant à la deuxième objection contre les expériences de Colaço et de ses collègues : la réflexion stimulée par ces expériences ne correspond pas à la réflexion dont font preuve les philosophes. Les philosophes ne se contentent pas de penser davantage à un problème et à une question ; ils comparent des cas différents, des expériences de pensée divergentes, et cherchent à trouver un équilibre entre différentes conceptions rivales. Cette objection possède une certaine valeur. Cependant, comme le mettent en avant Colaço et ses collègues, elle pose un problème radical, en tant qu'il ne semble pas toujours évident que les philosophes procèdent systématiquement par considération du plus grand nombre de cas possibles. Il n'est en ce sens pas évident que la considération d'une réflexion large parvienne véritablement à caractériser l'activité philosophique. Dans un certain nombre de cas, en effet, les expériences de pensée ne font pas appel à une comparaison d'une multiplicité de cas ; il s'agit uniquement de mettre en avant un cas particulier, qui est censé exclure de façon décisive une conception philosophique particulière. Par exemple, considérer un cas de Gettier est censé être en soi suffisant pour exclure l'idée que la connaissance puisse être définie comme une croyance vraie et justifiée.

Ces deux premières objections à l'expérience de Colaço et de ses collègues sont donc intéressantes, mais non concluantes. Cependant, un troisième problème, beaucoup plus grave, se pose pour l'idée que la réflexion ne changerait en rien l'activité philosophique : il existe bien des expériences montrant que la réflexion conduit les gens à changer d'avis, même dans le cas d'expériences de pensée utilisées par les philosophes. Deux expériences sont sur ce point particulièrement pertinentes pour notre propos. La première concerne le cas du dilemme du tramway, où il a été prouvé de façon claire qu'une plus grande réflexion conduit à des jugements davantage conformes à une éthique utilitariste. Par exemple, dans une série d'expériences publiées en 2015, similaires à celles discutées plus haut, Schwitzgebel et Cushman ont proposé à des philosophes et des non-philosophes de répondre à une série de dilemmes similaires au

¹⁵³ D. COLAÇO, M. KNEER, J. ALEXANDER et E. MACHERY, « On second thought », *op. cit.*

dilemme du tramway¹⁵⁴. Les participants étaient assignés ou bien à une condition visant à favoriser la réflexion, ou bien à une condition de contrôle. La condition « réflexion » consistait à instruire les participants de réfléchir systématiquement à leurs réponses, à prendre en compte des scénarios alternatifs à ceux proposés, et imposait un temps minimal nécessaire avant de fournir des réponses. Les auteurs ont découvert qu'imposer ces conditions aux participants a changé leurs réponses de façon assez nette, dans une direction plus conforme à une philosophie utilitariste. De façon intéressante, l'effet obtenu était similaire chez les philosophes et les non-philosophes. D'autres types d'expériences confirment qu'un surcroît de réflexion conduit les participants à adopter une attitude davantage conforme à une perspective utilitariste dans le cas de dilemmes sacrificiels ; par exemple, plusieurs études ont trouvé une corrélation positive entre le fait d'adopter des réponses conformes à une perspective utilitariste et le fait d'obtenir un score positif au *Cognitive Reflection Test*, une tâche courte visant à mesurer la réflexivité ou la capacité analytique des individus¹⁵⁵.

Une deuxième expérience a récemment tenté d'évaluer dans quelle mesure l'exposition à des contre-arguments pouvait changer les points de vue des individus. Il s'agissait d'évaluer l'importance d'un type particulier de réflexion, à savoir la manière dont les individus réagissent à des opinions argumentées différentes des leurs. Dans leur article, Matthew L. Stanley et ses collègues ont utilisé de multiples types de dilemmes moraux différents¹⁵⁶. Par simplicité, et parce que les différentes expériences ont donné des résultats concordants, je me concentrerai sur l'étude 1b, qui a été testée sur plus de 2000 participants. Dans l'étude 1b, trois types de dilemmes moraux classiques ont été utilisés ; par exemple, un des trois types correspondait au cas de la transplantation d'organes, qui implique d'avoir un choix à faire entre sauver un patient, ou le laisser mourir et se servir de ses organes pour sauver cinq autres personnes. Dans la version particulière utilisée par les auteurs, la personne en question était une personne sans abri, dont la vie n'était pas particulièrement en danger, et les cinq personnes étaient une famille en danger de mort à la suite d'un accident de voiture. Après avoir vu la description de chaque dilemme, les participants devaient s'imaginer être à la place de la personne devant effectuer le choix moral, et devaient indiquer le choix qu'ils feraient. Par exemple, dans le cas du don

¹⁵⁴ E. SCHWITZGEBEL et F. CUSHMAN, « Philosophers' biased judgments persist despite training, expertise and reflection », *op. cit.*

¹⁵⁵ Indrajeet PATIL, Micaela ZUCHELLI, Wouter KOOL, Stephanie CAMPBELL, Federico FORNASIER, Marta CALO, Giorgia SILANI, Mina CIKARA et Fiery Andrews CUSHMAN, « Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures », *manuscrit non publié*.

¹⁵⁶ Matthew L. STANLEY, Ashley M. DOUGHERTY, Brenda W. YANG, Paul HENNE et Felipe DE BRIGARD, « Reasons probably won't change your mind: The role of reasons in revising moral decisions », *Journal of Experimental Psychology: General*, 2018, vol. 147, n° 7, p. 962-987.

d'organes, ils devaient indiquer s'ils choisiraient ou non de faire la transplantation. Après avoir effectué ce choix, les participants étaient confrontés à une série de six arguments. Dans une condition, les participants lisaient des arguments en faveur de la position opposée à la leur ; dans l'autre condition, ils lisaient des arguments en faveur du choix qu'ils avaient effectué. Par exemple, dans le cas de la transplantation d'organes, les participants pouvaient lire des arguments du type : « La personne sans abri n'a pas consenti à la transplantation de ses organes. » contre la transplantation, ou des arguments du type : « Effectuer la transplantation d'organes sauverait le plus grand nombre de vies. » en faveur de la transplantation. Après avoir vu la série de six arguments, les participants se sont vus proposer la possibilité de changer leur choix. Dans le cas des participants ayant vu des raisons opposées à leur choix, selon les types de dilemmes, entre 8 et 12% des participants ont changé leur choix ; dans le cas des participants ayant vu des raisons favorables à leur choix, il n'y a jamais eu plus de 2% de participants qui ont changé de décision. L'impact de la condition était clair, et hautement significatif.

L'interprétation de ces résultats est délicate. D'un côté, la condition a manifestement eu un effet, et beaucoup plus de participants ont changé d'avis s'ils ont été exposés à des arguments opposés. D'un autre côté, la proportion d'individus ayant modifié leur décision est restée très faible : pas plus de 12% des participants ont changé leur avis. C'est ce dernier aspect que les auteurs ont mis en avant dans leur article, qui s'intitule en effet « Les raisons ne vont probablement pas vous faire changer d'avis ». Les auteurs insistent donc sur le fait que la grande majorité des participants sont restés constants dans leur choix, ce qui indique que la confrontation à des opinions opposées n'aura vraisemblablement un impact que sur une toute petite minorité d'individus. Les auteurs concluent que leurs résultats renforcent donc plutôt le modèle d'individus faisant leur choix pour des raisons intuitives, et pour lesquels les raisons sont davantage un moyen de justifier une décision prise de façon indépendante.

Une telle conclusion est contestable, pour une raison très simple : il est difficile d'établir une proportion d'individus qui « devraient », selon un modèle d'individus rationnels, répondre à la présentation d'opinions opposées. En l'absence d'un modèle précis, il est difficile de dire dans quelle mesure un chiffre de 12% d'individus changeant leur avis constitue une proportion importante ou faible. En effet, on peut supposer que des individus ont pris leur première décision après une considération du pour et du contre en faveur de chaque possibilité. Si tous les individus changeaient d'avis après une présentation d'arguments opposés, ce serait un signe d'irrationalité beaucoup plus fort que la relative constance découverte par les expérimentateurs. Un décideur rationnel est supposé être relativement constant dans ses décisions ; même si l'on peut s'attendre à ce que des arguments opposés qu'il n'avait pas considérés le fassent parfois

changer d'avis, il n'est pas évident que la considération d'arguments opposés doive le faire changer d'avis très souvent. Un impact modéré de la considération d'arguments opposés est donc compatible avec un rôle important de la réflexion, et est en tout cas incompatible avec des théories niant absolument tout impact à la considération rationnelle d'arguments.

Justifier le rôle du philosophe, sans le considérer comme un expert

Les précédentes sections ont établi plusieurs faits. Tout d'abord, les philosophes sont susceptibles d'être victimes des mêmes types de biais psychologiques que les non-philosophes. Ensuite, il n'est pas toujours évident qu'une plus grande réflexion conduise à un changement de jugement dans la considération de problèmes philosophiques. Plus précisément, une plus grande réflexion semble n'avoir aucun impact dans la considération de problèmes philosophiques particuliers. Cependant, ces résultats ne constituent pas une pure justification du programme négatif en philosophie expérimentale, et ne remettent pas complètement en cause l'usage d'expériences de pensée. Tout d'abord, la réflexion semble bien avoir un impact sur le jugement philosophique, au moins dans certains cas. Le problème n'est donc pas que la réflexion n'a aucun impact ; le problème est que la réflexion a un impact bien moindre que ce à quoi l'on pourrait s'attendre. Il y a une forte chance que les individus restent fixés dans les décisions qu'ils ont prises à première vue. En tant que tel, ce résultat doit donc inciter à une forte prudence philosophique.

La conclusion à prendre n'est donc pas que l'activité philosophique est complètement vaine, ou qu'il n'existe à coup sûr aucune expertise philosophique. Cependant, il est évident que, si expertise philosophique il y a, elle ne peut être considérée comme suffisamment forte pour disqualifier toute opinion contraire. La considération de ce qui est pris en compte par les individus doit amener à une remise en cause potentielle du philosophe. L'expertise philosophique ne semble pas telle que le philosophe puisse négliger la considération de données expérimentales, notamment si ces données montrent que l'option préférée par la majorité d'une population est une option opposée à celle préférée par le philosophe en question. La confrontation avec les données expérimentales semble importante. Si les intuitions du philosophe sont en accord avec celles d'une majorité des personnes interrogées, l'on peut considérer qu'il s'agit d'un premier signe que les opinions du philosophe ne sont pas en soi absurdes. Cependant, en cas de désaccord clair, il devient important de se demander pourquoi

il existe un écart entre les opinions du philosophe et celles d'une grande partie de la population. Est-ce parce que les participants sont biaisés par des facteurs émotionnels ou non-pertinents ? Est-ce parce que le philosophe lui-même serait biaisé ? Est-ce dû au fait que les participants ont mal compris le problème philosophique ? Si aucune raison claire ne se présente pour expliquer la divergence, le philosophe doit au minimum considérer cet écart comme une opportunité pour prendre de la distance par rapport à ses propres intuitions. La considération d'un désaccord philosophique peut donc être un moyen de renforcer la réflexivité de l'activité du philosophe.

Cette défense d'une confrontation des pensées du philosophe à l'opinion courante s'est jusqu'à présent faite sur la base d'un examen des biais de raisonnement inhérent à la considération d'expériences de pensée. Cependant, il existe une deuxième voie pour justifier l'utilité des idées courantes : on peut mettre en avant le fait qu'elles possèdent un type de sagesse. C'est vers cette deuxième voie que nous nous tournons maintenant.

Pourquoi l'opinion publique est plus sage que les philosophes (1) : la sagesse des foules

Jusqu'à présent, nous avons examiné des raisons en faveur de l'idée que le philosophe posséderait une plus grande expertise que l'opinion courante, et qui pourraient donc justifier le droit du philosophe à négliger les données expérimentales. Cependant, il est important de considérer deux raisons qui pourraient donner un résultat inverse, et qui accorderaient une plus grande compétence à l'opinion publique plutôt qu'au philosophe. Ces deux raisons sont le phénomène de sagesse des foules, d'une part, et le phénomène d'évolution des normes sociales, d'autre part. Ces deux facteurs pourraient justifier le fait d'accorder un bénéfice du doute à l'opinion publique, et donc de rejeter à première vue l'opinion du philosophe, en cas de désaccord entre les deux.

Le phénomène de la sagesse des foules a une longue histoire, en ce qu'il remonte au moins aux travaux de Francis Galton à la fin du XIXe siècle. L'idée de la sagesse des foules renvoie à des situations où l'on demande à un groupe d'individus d'estimer une quantité ou la probabilité d'un événement. L'idée de « sagesse » correspond au fait que la moyenne des réponses de ce groupe d'individus peut s'avérer extrêmement précise, ou, du moins, beaucoup plus précise que prendre la réponse d'un individu au hasard. En d'autres termes, la moyenne des jugements est souvent beaucoup plus exacte que la plupart des jugements individuels. Parmi

des exemples classiques, on peut citer le fait que prendre la moyenne des réponses peut être un excellent prédicteur de la température d'une pièce, du nombre de bonbons dans une jarre, ou du poids d'un bœuf¹⁵⁷. Dans le cas de la température de la pièce, étudiée par Knight en 1921, la moyenne des réponses s'approchait davantage de la vérité que 80% des réponses des participants, et dans le cas des bonbons, la moyenne était plus exacte que 98% des estimations des participants. Le concept de sagesse des foules a aussi été récemment mobilisé dans un usage politique afin de justifier les procédés de décision collective, y compris dans un cadre démocratique¹⁵⁸.

Naturellement, l'hypothèse d'une sagesse des foules ne constitue pas une justification parfaite de l'opinion collective. Tout d'abord, il est évident que les cas où l'idée de sagesse des foules marche le mieux sont des cas qui impliquent une tâche relativement simple (comme donner le poids d'un bœuf), où il existe une réponse exacte, et où l'on peut supposer que les erreurs faites par les différents individus ne sont pas biaisées dans une direction particulière. Ainsi, dans un article critique vis-à-vis de la possibilité d'une sagesse des foules étendue à tout contexte possible, Joe Simmons et ses collègues mettent en avant quatre conditions qui doivent être réunies pour qu'une sagesse des foules puisse avoir lieu. « En somme, les partisans de la sagesse des foules prédisent que les foules seront sages quand les juges 1) connaissent le sujet traité, 2) désirent donner une réponse exacte, 3) sont indépendants et 4) sont divers¹⁵⁹. » Les auteurs ajoutent, de plus, qu'indépendamment de ces quatre conditions, la sagesse des foules peut échouer si les participants sont tous biaisés dans une même direction.

Si l'on veut appliquer l'idée de sagesse des foules aux questions de philosophie morale, il faut bien avouer qu'il n'est en rien évident que ces différentes conditions nécessaires à l'action positive de la sagesse des foules soient réunies dans le cas de l'évaluation des problèmes philosophiques.

Il est aisé de voir pourquoi les conditions de la sagesse des foules ne sont pas parfaitement réunies dans l'examen des questions morales. Tout d'abord, il est difficile de dire dans quelle mesure les individus possèdent une bonne connaissance du sujet traité : avoir une connaissance de ce que constitue la justice est beaucoup plus dur à définir que savoir approximativement combien pèse un bœuf. Ensuite, il n'est pas évident que les différents participants interrogés

¹⁵⁷ Cf. Les articles classiques de Knight, Treynor et Galton, cités dans Joseph P. SIMMONS, Leif D. NELSON, Jeff GALAK et Shane FREDERICK, « Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds », *Journal of Consumer Research*, 2011, vol. 38, n° 1, p. 1-15.

¹⁵⁸ Hélène LANDEMORE et Jon ELSTER (dir.), *Collective wisdom: principles and mechanisms*, Cambridge, Cambridge Univ. Press, 2012.

¹⁵⁹ J. P. SIMMONS, L. D. NELSON, J. GALAK et S. FREDERICK, « Intuitive Biases in Choice versus Estimation », *op. cit.*

lors d'enquêtes soient capables de donner des réponses indépendantes et non-biaisées. En effet, les différents individus d'un même pays partagent une même culture, de mêmes institutions, et pour cela ont certainement tendance à posséder une même vision de la justice ou du mérite, qui n'est pas nécessairement conforme à ce qu'on pourrait imaginer comme une justice idéale¹⁶⁰.

Cependant, même si l'on admet que l'existence d'une sagesse des foules est douteuse dans le cas de la résolution des problèmes philosophiques, il est difficile de considérer que le philosophe disposerait d'un point de vue privilégié qui le mettrait à l'écart des facteurs qui peuvent remettre en question la sagesse collective. Reprenons en effet les deux difficultés qui se posent face à l'existence d'une sagesse des foules. Tout d'abord, s'il est difficile de définir en quoi peut consister une connaissance de ce que constitue la justice, il est évident que cette difficulté peut remettre en cause également l'idée qu'un philosophe pourrait être compétent pour traiter cette question. Ensuite, il est évident qu'un philosophe tend à partager la même culture que ses concitoyens ; dans ce cadre, un ensemble de préjugés ou de croyances partagées dans sa communauté tendront à l'influencer d'une manière similaire, ce qui fait que tout facteur culturel ayant une influence négative sur la sagesse des foules tendra à affecter le philosophe de la même manière.

Dans ce cadre, il est possible de considérer que la foule conserve un avantage vis-à-vis du philosophe, en tant qu'elle consiste toujours en l'accumulation d'opinions diverses sur un sujet. L'agrégation aboutira à la conclusion d'un comportement moyen, et limitera la possibilité que le jugement des foules soit un jugement extrême. Bien sûr, un jugement moyen n'est pas forcément un jugement juste, surtout dans le cadre où il existe des tendances systématiques au biais dans une direction particulière. Cependant, on peut quand même considérer que le filtrage des opinions extrêmes est un avantage, même dans un cadre où l'opinion moyenne n'est en rien garantie de vérité.

Pourquoi l'opinion publique est plus sage que les philosophes (2) : l'évolution de la morale

L'argument en faveur des foules n'est donc pas entièrement concluant par lui-même. Cependant, il est possible de doubler cet argument par un deuxième argument en faveur de

¹⁶⁰ Même si savoir en quoi consisterait une justice idéale peut être de toute façon problématique à définir.

l'idée que l'opinion commune serait plus proche de la vérité que l'opinion des philosophes. Cet argument, défendu de façon la plus nette par des auteurs comme Edmund Burke ou Friedrich Hayek, consiste à soutenir que l'opinion ou les pratiques publiques ont connu une évolution culturelle et sociale qui les a conduites à une certaine adéquation vis-à-vis de la réalité. Les deux auteurs doublent cette défense d'une évolution culturelle par une critique de la capacité d'innovation individuelle, et soutiennent qu'une reconstruction non-guidée par l'histoire sera vraisemblablement incapable de mener à une théorie en adéquation avec la réalité.

La critique avancée par Burke des innovations individuelles est particulièrement intéressante pour notre propos, parce qu'elle le conduit à une défense des préjugés communs¹⁶¹. Dans sa critique des révolutionnaires français, Burke défend l'idée que l'innovation faite à partir de rien, non basée sur les structures existantes, risque de tout compromettre parce qu'il est difficile de voir les défauts d'un système qu'on invente. Des idées abstraites sont dangereuses en tant qu'elles ne sont pas adaptées à la situation présente. Cette critique de Burke vise notamment les droits de l'homme abstraits inventés par les révolutionnaires français, qui souffrent d'un tel degré de généralité qu'ils ne sont pas capables de guider la pratique et les besoins présents. La comparaison avec le manque de plausibilité de théories philosophiques abstraites se fait naturellement.

Contre cette incapacité de l'innovation *de nihilo* à construire des idées ou des pratiques adaptées au réel, Burke défend la valeur d'une évolution culturelle progressive. Cette amélioration progressive des institutions se fonde notamment sur les avantages d'un tâtonnement empirique, c'est-à-dire sur la tentative d'améliorer marginalement les institutions existantes. Burke défend notamment l'idée qu'une transformation limitée a moins de chances qu'une innovation radicale de mener à de grands défauts, ce qui permet de maintenir les avantages des pratiques existantes.

Même si Burke se fait surtout défenseur des institutions britanniques, il met aussi en avant la valeur pratique des idées populaires, et notamment des préjugés. Voici le passage où il défend le plus clairement leur valeur :

« Beaucoup de nos penseurs, au lieu de mettre au rebut les préjugés communs, emploient toute leur sagacité à découvrir la sagesse cachée qu'ils renferment. S'ils parviennent à leur but, et rarement ils le manquent, ils estiment qu'il vaut mieux garder le préjugé avec ce qu'il contient de raison que de se défaire de l'enveloppe pour ne garder que la raison toute nue ; et cela parce qu'un préjugé donne à la raison qu'il contient le motif qui fait sa force agissante et l'attrait qui

¹⁶¹ Edmund BURKE, *Réflexions sur la Révolution en France*, Paris, Les Belles lettres, 2016.

assure sa permanence. En cas d'urgence le préjugé est toujours prêt à servir ; il a déjà déterminé l'esprit à ne s'écarter jamais de la voie de la sagesse et de la vertu, si bien qu'au moment de la décision, l'homme n'est pas abandonné à l'hésitation, travaillé par le doute et la perplexité. Le préjugé fait de la vertu une habitude, et non pas une suite d'actions isolées. Par le préjugé fondé en raison, le devoir entre dans la nature de l'homme¹⁶². »

Burke mêle deux idées ici pour défendre la valeur des préjugés, et pour mettre en avant le fait qu'il est préférable de conserver les préjugés plutôt que de chercher à les remplacer par une idée fondée en raison. La première raison en faveur des préjugés est le fait qu'ils contiennent vraisemblablement une part de vérité, façonnée par un processus d'évolution culturelle. La deuxième raison est que les préjugés sont déjà intégrés au sein de la population, ce qui fait qu'ils possèdent une valeur pratique évidente. Si l'on tentait de remplacer les préjugés, le remplacement resterait sûrement à un niveau purement cognitif, sans conduire à une modification des pratiques ; et s'il y avait une modification des pratiques, celles-ci risqueraient de rester à un niveau beaucoup moins intuitif et efficace que les préjugés ancrés. Pour prendre un exemple, dans la psychologie contemporaine, les théoriciens de la rationalité écologique, notamment Gerd Gigerenzer, ont défendu l'idée que fonder son raisonnement sur des heuristiques simples avait souvent de meilleures conséquences que de chercher à fonder son choix sur l'option la plus rationnelle possible, c'est-à-dire capable de maximiser les résultats voulus¹⁶³. En effet, dans un certain nombre de situations, chercher la maximisation de résultats est tout simplement trop coûteux, voire impossible d'un point de vue cognitif. Rester fidèle à des idées heuristiques, voire à des préjugés, peut finalement s'avérer adapté dans un certain nombre de situations.

Cette défense des préjugés et des opinions populaires peut être renforcée par la considération des arguments de Friedrich Hayek. Même s'il existe globalement un accord fort entre les analyses de Burke et celles de Hayek, certains points développés par ce dernier dans *Droit, législation et liberté* nous seront particulièrement utiles, en tant qu'ils permettent de renforcer l'idée que les conceptions courantes peuvent être supérieures à celles d'un philosophe isolé¹⁶⁴.

Tout d'abord, Hayek renforce l'idée de Burke contre les avantages d'une innovation radicale en se fondant sur une étude de la distribution de la connaissance au sein d'une société moderne. Pour qu'une reconstruction radicale des règles sociales soit possible, il faudrait, selon Hayek,

¹⁶² Cité dans l'anthologie de Pierre MANENT (dir.), *Les libéraux*. 2, Paris, Hachette, 1986, p. 23.

¹⁶³ Gerd GIGERENZER et Peter M. TODD (dir.), *Simple heuristics that make us smart*, Oxford, Oxford Univ. Press, 2001.

¹⁶⁴ F. A. HAYEK, *Law, Legislation and Liberty*, op. cit.

qu'un individu puisse disposer d'un ensemble de connaissances supérieures à celles déjà existantes au sein de la société. Or, si une telle situation est envisageable dans des sociétés à petite échelle, elle semble impossible dans une société moderne où des millions d'individus interagissent. Ainsi que l'écrit Hayek : « L'économie a depuis longtemps insisté sur la « division du travail » impliquée dans une telle situation. Mais elle a beaucoup moins insisté sur la fragmentation de la connaissance, sur le fait que chaque membre de la société ne peut posséder qu'une petite partie de la connaissance totale, et sur le fait que chaque individu est donc ignorant de la plupart des faits sur lesquels se fonde le fonctionnement de la société. Et cependant, c'est la mise en commun de bien plus de connaissances que ce que n'importe qui peut posséder, et donc le fait que chacun se situe dans une structure cohérente dont la plupart des déterminants lui sont inconnus, qui constitue le trait caractéristique de toutes les civilisations avancées¹⁶⁵. » Hayek met donc en avant le fait que la connaissance collective est bien supérieure à la connaissance de n'importe qui, et même des plus grands experts. En conséquence, seule une structure sociale qui maximise la possibilité de l'utilisation de ces connaissances peut être une structure viable.

En plus de cette justification d'une connaissance distribuée, les théories mises en avant par Hayek permettent de résoudre une difficulté dans la défense des idées courantes. En effet, on pourrait objecter qu'il n'est pas très raisonnable de se fonder sur l'opinion commune, en tant qu'il est très peu probable que la plupart des individus soit capable de reconstruire rationnellement et explicitement une théorie de la justice sociale. De fait, ce n'est pas ce que cherchent à faire les théories du jugement moral : celles-ci cherchent la plupart du temps à reconstruire de manière explicite les jugements faits implicitement par les individus. Or, Hayek, dans sa théorie, cherche à mettre en avant pourquoi les individus peuvent avoir des jugements implicites parfaitement adaptés à leur situation, alors même qu'ils seraient tout à fait incapables de construire des théories explicites cohérentes.

Hayek met en effet en avant le fait qu'une structure efficace, basée sur l'évolution sociale, n'est possible que si les individus ont intériorisé des règles qui leur permettent de s'adapter à leur environnement actuel. Toujours selon Hayek, « l'homme est un animal qui est autant caractérisé par le fait de suivre des règles que par la quête de sens. Et il n'est pas seulement couronné de succès parce qu'il sait pourquoi il doit se conformer aux règles auxquelles il se conforme en pratique, ou même parce qu'il est capable d'édicter toutes ces règles, mais parce que sa pensée et son action sont gouvernées par des règles qui ont évolué par un processus de

¹⁶⁵ *Ibid.*, p. 14.

sélection dans la société dans laquelle il vit, et qui sont donc le produit de l'expérience de générations¹⁶⁶. » Hayek défend donc l'idée que des jugements implicites peuvent être parfaitement adaptés à l'environnement social, et que la connaissance commune peut être gigantesque sans jamais avoir à être consciemment pensée par des individus. Dans ce cadre, il est possible de défendre l'idée que les expériences psychologiques contribuent à mettre en avant une connaissance commune implicite qui reflète une adaptation à l'environnement. L'opinion commune, telle qu'elle est reconstituée dans le cadre d'expériences psychologiques, posséderait alors une supériorité épistémique face à la reconstruction solitaire de principes de vie commune exercée par le philosophe. Ce dernier, privé de la connaissance partagée, serait incapable de construire des principes adaptés à un environnement moderne trop complexe.

Je n'ai pas présenté les arguments évolutionnistes défendus par Burke et Hayek parce qu'ils sont irréfutables, ou nécessairement parfaitement convaincants. Le lecteur n'aura sans doute pas trop de difficultés à trouver des exemples historiques de préjugés ne contenant aucun contenu empirique, ou des exemples de pratiques collectives largement erronées. Cependant, même dans le domaine de l'histoire des préjugés, il est assez notable que l'évolution culturelle moderne a conduit à l'élimination et à la mise à l'écart des préjugés les plus dommageables, notamment dans le domaine du racisme et de l'antisémitisme. Des auteurs en sciences sociales comme Steven Pinker ont ainsi pu défendre l'idée que les derniers siècles ont vu un progrès net dans tous les domaines, y compris dans les pratiques morales et sociales¹⁶⁷. Même si l'on considère que les théories de l'évolution culturelle défendues par Hayek et Burke ne fonctionnent pas toujours, elles fournissent cependant un argument à première vue pour prendre les opinions populaires au sérieux. Celles-ci contiennent bien un degré d'information et de connaissance qui n'est pas forcément contenu dans les opinions privées des philosophes. Si croire aveuglément à n'importe quelle théorie populaire serait évidemment une grave erreur, exiger des philosophes qu'ils se confrontent aux opinions répandues, en prenant au sérieux le fait que celles-ci ont pu faire l'objet d'une sagesse des foules et d'une évolution culturelle, semble absolument nécessaire.

Les arguments précédemment cités en faveur de la sagesse des foules, et en faveur de l'idée que les croyances communes ont été le résultat d'un processus d'évolution culturelle qui a probablement conduit à la sélection d'idées correctes, peut sembler douteux, voire même absurde. Je me limiterai cependant ici à faire appel au lecteur : s'il avait le choix entre défendre les idées les plus répandues au sein de la population, ou défendre les idées défendues par un

¹⁶⁶ *Ibid.*, p. 12.

¹⁶⁷ Steven PINKER, *The better angels of our nature*, London, Penguin, 2012.

philosophe qui serait tiré au hasard au sein de la population universitaire, quel serait son choix ? Il est important de rappeler ici qu'un philosophe tiré au hasard aurait une forte chance de défendre une théorie utilitariste, un matérialisme éliminativiste, un panpsychisme ; sur le point de vue moral ou politique, il aurait des chances (minimes, certes) de défendre l'infanticide comme choix moral acceptable. Une grande partie des philosophes défendent certes des positions qui sont beaucoup plus consensuelles. Mais le risque d'avoir à défendre des positions à première vue absurdes devrait jouer son rôle dans la justification du sens commun comme source de positions philosophiques *a priori* respectables.

Conclusion : La philosophie expérimentale comme surcroît de réflexivité

Ce chapitre a présenté une série de raisons pour réfuter l'idée qu'il existerait un gouffre d'expertise entre les philosophes et les non-spécialistes. Bien sûr, les philosophes ont beaucoup plus réfléchi aux problèmes philosophiques fondamentaux. Cependant, il est assez douteux que cette expérience beaucoup plus grande rendrait complètement stérile l'examen des idées courantes relatives aux questions philosophiques. En effet, comme nous l'avons vu, les philosophes, et les experts en général, restent soumis à un certain nombre de biais, même lorsqu'ils considèrent des cas qui font partie de leur domaine d'expertise. De plus, il existe des données contradictoires sur l'impact de la réflexivité sur les réponses apportées aux problèmes philosophiques, de telle sorte qu'il n'est en rien évident qu'un surcroît de réflexion changera forcément l'opinion de quelqu'un sur une question. Enfin, les opinions courantes possèdent deux vertus. Tout d'abord, en tant qu'opinions agrégées, elles bénéficient de l'effet dit de « sagesse des foules », qui implique que l'agrégation de différentes options conduit à l'annulation d'erreurs allant dans des directions opposées, et peut permettre d'améliorer la qualité de la réponse. Ensuite, les opinions courantes sont le résultat d'un processus d'évolution culturelle de plusieurs millénaires, au cours duquel certaines des opinions les plus aberrantes d'un point de vue moral ont pu être largement éliminées.

Bien sûr, il serait absurde de se fier aveuglement aux opinions courantes. Personne ne douterait qu'elles peuvent être très largement dans l'erreur sur des questions de moralité. Mais le philosophe ne se trouve pas dans une position épistémique si privilégiée qu'il puisse se permettre de les mettre complètement de côté. C'est cette position minimale que je demande au

lecteur d'accepter, à défaut d'une position maximale qui considérerait que les foules sont forcément plus sages que les philosophes.

Maintenant que ces présupposés méthodologiques ont été acceptés, il est temps de passer à une défense complète du mérite. Nous allons maintenant nous attaquer à l'examen des quatre principales justifications du mérite.

Troisième partie
Justifier le mérite

Chapitre 6

Justifier le mérite : Trois premières approches

Les 150 dernières années ont vu la mise en œuvre de multiples théories extrêmement critiques vis-à-vis de l'idée de mérite. Alors que les théories classiques du mérite, depuis Platon et Aristote jusqu'à Adam Smith ou Kant, considéraient que la validité des jugements de mérite était une évidence, les théories modernes ont systématiquement cherché à démontrer que la valeur du mérite était une illusion. Au XIXe siècle, les théories utilitaristes défendues par John Stuart Mill ou Henry Sidgwick ont cherché à démontrer que les intuitions liées au mérite devaient être réinterprétées comme des intuitions relatives au moyen de servir au mieux l'utilité collective. Au XXe siècle, les critiques, à la suite notamment de John Rawls, ont cherché à montrer que le mérite n'était pas justifiable par lui-même, mais uniquement comme un moyen pour améliorer la situation des plus défavorisés. Les théories ont été rares qui ont accordé au mérite une place fondamentale, et qui ont considéré l'idée comme pleinement justifiée.

A l'inverse de ces doutes modernes, je mettrai ici en avant la multiplicité de justifications possibles du mérite. Dans chaque cas, je montrerai qu'une justification est non seulement possible, mais également pleinement valide. En d'autres termes, loin d'être une idée privée de justification normative, il semble possible de défendre le mérite à partir d'une pluralité de points de vue. Dans tous ces cas, nous reviendrons donc à l'idée classique d'une évidence de la validité des jugements de mérite. De fait, nos arguments se fonderont en grande partie sur ces théories classiques, et notamment sur la défense magistrale du mérite formulée par Adam Smith dans sa *Théorie des sentiments moraux*.

Comme nous le verrons plus en détails, il existe quatre justifications majeures du mérite. Une de ces justifications est purement instrumentale : elle consiste à justifier le mérite à partir de l'utilité générée par la récompense d'une action positive et par la punition des actions néfastes. La première section de ce chapitre détaillera davantage ce premier type de justification, et montrera pourquoi il ne s'agit pas d'une défense adéquate de l'idée de mérite. Même si cette justification instrumentale est en soi une bonne nouvelle pour les défenseurs du

mérite, il ne s'agit évidemment pas d'une justification montrant que le mérite possède une valeur intrinsèque. Je montrerai notamment pourquoi quelqu'un attaché à une justification purement instrumentale du mérite devrait accepter de nombreuses situations en violation flagrante avec nos idées courantes de mérite.

Pour cette raison, les sections suivantes examineront tour à tour trois autres types de justification, qui accordent une importance beaucoup plus grande à l'idée de mérite. Un deuxième type de justification accorde au mérite une valeur intrinsèque. Selon cette deuxième justification, il est bon de récompenser le mérite parce qu'il s'agit d'une intuition morale fondamentale. Selon cette perspective, le mérite serait justifié parce qu'il semble intuitivement bon. Une telle justification basée sur de pures intuitions pourra sembler scandaleuse au lecteur. Nous montrerons pourtant qu'il s'agit d'une démonstration parfaitement valide de la valeur *prima facie* du mérite.

Dans un troisième temps, nous examinerons l'idée que le mérite sert à concilier deux valeurs souvent opposées : le bien commun et le bien individuel. Le mérite est alors considéré comme l'instrument nécessaire pour assurer la congruence entre deux valeurs morales fondamentales.

Enfin, dans un quatrième temps, nous examinerons une quatrième justification du mérite : l'idée que le mérite sert à communiquer l'approbation ou la désapprobation sociale à l'individu récompensé ou puni. Cette perspective a été fortement mobilisée dans la théorie de la peine contemporaine, afin de trouver une voie entre justifications purement rétributives et justifications purement dissuasives. Le chapitre 7 constituera un examen complet de cette idée, dans une perspective à la fois empirique et normative.

La justification instrumentale du mérite : examen des théories institutionnelles du mérite

La première grande justification du mérite consiste en une justification instrumentale en termes d'efficacité. Selon cette conception, mettre en place un système qui récompense des individus ayant effectué une action positive fournit une incitation à bien agir par la suite. En conséquence, cette récompense conduit à un plus grand nombre d'actions positives ayant été effectuées. Il peut en découler un plus grand bien-être général. Par exemple, récompenser un individu qui a bien fait son travail le conduira sans doute à effectuer de nouveau un bon travail

par la suite, ce qui augmentera la production. En conséquence, il est bon de récompenser cette personne, parce que cela conduira à une plus grande efficacité à l'avenir.

Pour prendre un exemple concret, on peut considérer que payer des médecins davantage que la moyenne conduit à un plus grand bien-être, étant donné que cela fournit une incitation à faire de longues études et à procurer un service indispensable à la société. Si les médecins étaient payés exactement le même salaire que les autres personnes, alors on peut supposer qu'un nombre bien moins important de personnes feraient des études de médecine, ou choisiraient de devenir médecin. De plus, on peut considérer qu'il s'agit d'un métier difficile et important, et qu'il est nécessaire pour le bien-être collectif d'attirer vers cette carrière les personnes les plus intelligentes de la société. Dans ce cadre, récompenser les médecins selon leur mérite contribue à un bien-être social plus important, et peut donc être justifié du fait des incitations ainsi produites.

Cette justification du mérite se fait sur une pure appréciation des conséquences, sans que la récompense du mérite soit considérée comme un bien en soi. Il est évident qu'une telle justification offre une défense purement instrumentale du mérite. Cette focalisation sur les conséquences de la récompense du mérite veut dire qu'il peut y avoir plusieurs types de théories très différentes accordant une valeur instrumentale au mérite. Par exemple, Henry Sidgwick, dans *The Methods of Ethics*, tout comme John Rawls dans sa *Théorie de la justice*, ont proposé une justification instrumentale du mérite tout à fait similaire, alors même que chaque théorie considèrerait que les objectifs finaux servis par la récompense du mérite étaient très différents¹⁶⁸. Dans *The Methods of Ethics*, Sidgwick considère que la récompense du mérite peut recevoir une justification utilitariste, en tant que récompenser une activité productive peut être un moyen de servir le bien-être du plus grand nombre sur le long terme. Dans sa *Théorie de la justice*, Rawls défend l'idée que la récompense du mérite peut avoir une place dans une société juste, mais uniquement si récompenser des individus pour leurs bonnes actions sert à améliorer la situation des plus défavorisés. Dans les deux cas, la justification est instrumentale, mais les buts finaux sont différents : le but final est l'utilité collective chez Sidgwick, et l'amélioration du sort des plus défavorisés chez Rawls.

Considérer que les conséquences de la récompense du mérite sont souvent positives, et servent généralement le bien-être collectif, est une idée qui est en soi extrêmement consensuelle. De manière générale, des considérations de ce genre servent à renforcer la légitimité d'une distribution qui se ferait conformément au mérite. Cependant, cette idée d'une

¹⁶⁸ H. SIDGWICK, *The Methods of Ethics*, op. cit. ; J. RAWLS, *Théorie de la justice*, op. cit.

efficacité de la récompense du mérite cache une certaine ambiguïté. En effet, on peut effectuer deux interprétations possibles de l'idée que la récompense du mérite sert l'intérêt général ou l'efficacité collective. Une de ces interprétations est généralement décrite comme étant pré-institutionnelle, et la deuxième est généralement décrite comme théorie institutionnelle du mérite. Et, comme nous allons le voir bientôt, ces deux interprétations aboutissent à des conclusions opposées quant à la place du mérite dans une théorie de la justice sociale : les théories pré-institutionnelles lui accordent une place authentique, tandis que les théories institutionnelles le considèrent comme un simple instrument au service de l'efficacité collective. En d'autres mots, il s'agit de savoir si l'efficacité est considérée comme étant la seule raison d'être du mérite, comme le défendent les théories institutionnelles du mérite, ou bien si le mérite possède une raison d'être et une justification indépendante, comme le défendent les théories pré-institutionnelles du mérite.

Selon une première interprétation, qui correspond aux théories pré-institutionnelles du mérite, l'idée de mérite est définie de façon indépendante de celle de savoir si récompenser les actions « méritantes » conduit à augmenter le bien-être général ou l'efficacité collective. Par exemple, on peut considérer que le mérite de quelqu'un dépend de la productivité de son action, selon les critères que nous avons dégagés dans le chapitre 2. Dans ce cadre, il n'y a aucune nécessité logique à ce que récompenser les actions méritantes serve l'efficacité collective ; il s'agit, tout au plus, d'un simple constat empirique. Par exemple, si l'on considère un individu qui effectue des actions pour le bien commun, mais qui n'attend aucune récompense en retour et le fait simplement parce qu'il considère qu'il s'agit d'une bonne action, alors par définition le fait de le récompenser n'influera pas sur son comportement. Dans le cas particulier des individus effectuant des actions indépendamment de la possibilité d'une récompense extérieure, récompenser le mérite ne sert aucune efficacité collective. Si tous les individus étaient motivés de façon intrinsèque à effectuer de bonnes actions, alors la récompense du mérite n'aurait aucune fonction d'utilité collective.

Bien sûr, l'hypothèse que les individus sont uniquement motivés de façon intrinsèque n'est pas en soi très probable. Dans la plupart des cas, nous sommes bien sûr également sensibles aux incitations extérieures. Si un travail possède une récompense, nous serons probablement davantage motivés pour l'effectuer. En règle générale, il semble empiriquement plausible que la récompense des actions positives pour la collectivité conduit à une augmentation du nombre de ces actions. Il existe donc une corrélation entre la récompense d'un mérite définie de façon pré-institutionnelle et l'efficacité collective. Mais cette corrélation n'est pas une nécessité, ou ne découle pas de la façon dont nous avons défini le mérite.

Selon les théories institutionnelles du mérite, au contraire, le « mérite » d'un individu ne doit pas être défini de façon indépendante de la question de savoir dans quelle mesure le récompenser servirait l'intérêt collectif. Il s'agira ici de montrer qu'une théorie purement institutionnelle du mérite ne correspond pas aux intuitions courantes au sujet du mérite. Si la seule justification possible du mérite était instrumentale, il faudrait en conclure que l'idée de mérite n'a pas sa place parmi les principes fondamentaux de justice.

La nécessité du rejet des théories institutionnelles du mérite

Nous avons vu dans le chapitre 1 les principales caractéristiques d'une théorie institutionnelle du mérite : savoir quelles actions doivent être récompensées est défini de façon indépendante de savoir si cette action est bonne ou non. Selon une théorie institutionnelle du mérite, une action ne doit pas être récompensée parce qu'elle est bonne. Cette action doit être récompensée parce que, si elle n'avait pas été récompensée, elle n'aurait pas été effectuée et le bien-être collectif en aurait pâti. Une action n'est source de mérite que si la récompense sert d'incitation à l'effectuer.

A partir de là, il est facile à voir que nous nous éloignons durablement de l'idée d'une valeur intrinsèque du mérite. Mais il faut aller plus loin : la théorie institutionnelle du mérite n'est pas simplement différente des idées courantes relatives au mérite ; elle leur est directement opposée. Adopter une théorie institutionnelle du mérite est une rupture radicale avec les idées communes.

Pour voir en quoi la théorie rawlsienne, et en général les théories institutionnelles du mérite, sont en rupture avec l'opinion populaire, on peut prendre des exemples simples d'individus effectuant des actions bonnes de façon désintéressée. Dans son livre *On Justice*, J. R. Lucas prend les exemples suivants comme objections à la théorie rawlsienne :

« Une nation reconnaissante offrit au duc de Malborough un palais à Blenheim. Sir Alexander Fleming et Lord Florey ont été couronnés d'honneurs et de prix pour leur découverte de la pénicilline. L'inventeur de la photocopieuse *Xerox* a fait fortune. Rawls devrait dire qu'il s'agit d'une injustice, sauf si l'on pouvait montrer qu'ils étaient uniquement intéressés par l'argent, et qu'ils n'auraient pas délivré ces biens sauf en l'échange d'un paiement. La plupart des gens rejetteraient cette conclusion du fait de sa perversité. Si, poussé par l'amour du genre humain, je travaille pour inventer un nouveau traitement pour le rhume, et si après avoir réussi j'admets que je n'aurais pas été dissuadé dans mes efforts si j'avais su que je n'obtiendrai aucun

avantage financier, alors je ne devrais pas être payé un centime ; mais si je négocie à chaque étape pour obtenir le revenu financier le plus important, et si je menace régulièrement, si l'on ne cède pas à mes demandes, d'aller faire de la recherche sur la guerre bactériologique pour les Russes, ou bien sur la production de produits de beauté pour une firme cosmétique, alors je mérite tous les milliers de livres sterling que je peux extorquer à mes concitoyens¹⁶⁹. »

Dans cette citation, Lucas simplifie les débats, mais est, fondamentalement, parfaitement correct. Disons un instant pourquoi les arguments de Lucas représentent une simplification de la position rawlsienne. Rawls ne dit évidemment pas qu'il faut évaluer au cas par cas les positions de chaque individu pour connaître leurs motivations. Il défend l'idée qu'il faut établir des règles générales pour distribuer les biens au sein de la société, et il n'est pas évident que ces règles, dans une société rawlsienne idéale, consisteraient à étudier au cas par cas la psychologie des entrepreneurs, des scientifiques, ou de l'ensemble de la population.

Cependant, malgré cette simplification, l'analyse de Lucas touche en plein dans le mille. En effet, une analyse rawlsienne n'offre aucune objection de principe à ne pas récompenser les individus généreux, et à uniquement récompenser les personnes avares ou seulement intéressées par l'argent. S'il était possible d'identifier de façon fiable le caractère des individus, il serait ainsi envisageable de déterminer une grille de salaire, les personnes les plus pro-sociables recevant le moins d'argent, et les personnes les plus égoïstes recevant le plus. On pourrait imaginer, par exemple, le fait de faire passer des tests de personnalité aux individus avant qu'ils puissent se présenter sur le marché du travail¹⁷⁰.

Il est facile de trouver d'autres expériences de pensée allant dans le sens proposé par Lucas. On peut par exemple imaginer une communauté religieuse qui, du fait de ses convictions profondes, serait prête à œuvrer pour le bien commun indépendamment de toute récompense matérielle. Il semble un peu paradoxal de considérer que la justice exige de ne les récompenser que très faiblement, et moins que d'autres travailleurs effectuant une contribution similaire, afin de distribuer l'argent économisé aux plus défavorisés. L'idée qu'un travail égal implique un salaire égal semble être profondément ancré dans les intuitions ordinaires.

¹⁶⁹ J. R. LUCAS, *On Justice*, Oxford, Oxford University Press, 1980, p. 189.

¹⁷⁰ Evidemment, de telles propositions posent des problèmes évidents, puisque même les individus les plus généreux auraient alors des incitations à cacher leur véritable niveau de générosité. Mais là n'est pas la question : une théorie institutionnelle du mérite ne dispose d'aucun argument de principe contre une telle approche, mais uniquement des objections pratiques. De la même manière, un utilitariste dispose d'objections pratiques contre l'idée de tuer quelqu'un pour voler ses organes et sauver par là davantage de vies humaines, mais ne dispose d'aucune objection de principe. Dans les deux cas, les expériences de pensée permettent de révéler le gouffre entre les intuitions ordinaires et ces systèmes philosophiques.

Ces arguments ne démontrent pas que l'efficacité n'est pas une justification possible de l'idée de mérite. Mais ils montrent bien que l'efficacité ne saurait être la seule justification invoquée, du moins si l'on ne désire pas créer une rupture brutale avec les intuitions ordinaires sur le sujet. Si l'on considère que la récompense des plus « méritants » est uniquement un moyen de promouvoir d'autres objectifs, sans posséder aucune valeur en soi, alors l'on entre dans un terrain où la définition de qui est « méritant » sera forcément en rupture avec les intuitions ordinaires sur le sujet. Une conception purement instrumentale du mérite n'a aucune raison de déterminer le mérite de quelqu'un à partir de ce que cette personne contribue à la communauté, ou à partir de la productivité du travail de cet individu. Une conception instrumentale du mérite doit définir le « mérite » de quelqu'un à partir de la production supplémentaire, ou du bien-être général supplémentaire, créé par la récompense de cet individu. Mais dans ce cas, on aboutit à une situation où les personnes qui contribuent au bien-être collectif par pur amour de leur prochain ne disposent d'aucun « mérite », parce que récompenser leur action ne sert absolument à rien. Une théorie institutionnelle du mérite ne dispose d'aucune conception qui se rapprocherait des intuitions ordinaires ; elle ne dispose que d'une conception du mérite en termes d'incitations.

Une justification instrumentale du mérite peut donc servir de soutien à une théorie du mérite, mais ne peut servir d'une justification à elle seule. Il est donc temps de passer à une deuxième justification du mérite, fondée sur son intuitivité.

La justification conservatrice du mérite : défendre le mérite comme valeur populaire

La deuxième justification du mérite que nous allons examiner peut être désignée comme « conservatrice », en tant qu'elle cherche à justifier le mérite du simple fait qu'il s'agit d'une idée morale déjà largement présente au sein de la population. En d'autres termes, il s'agit de considérer que le fait que l'idée du mérite est répandue constitue en soi une justification suffisante de cette idée. Le fait que cette justification est « conservatrice » ne possède ici aucun sens politique spécifique : l'idée d'égalité est aussi une idée très populaire dans les sociétés contemporaines, et il est donc possible de fournir une justification conservatrice de l'idée d'égalité en se basant sur le fait qu'il est bon de conserver les idées morales établies.

Nous avons vu dans le chapitre 3 que l'idée de mérite est effectivement très populaire dans les sociétés contemporaines. On trouve un très fort soutien à l'idée de récompense du mérite, à la fois quand l'idée est exprimée de façon abstraite et concrète, et à la fois dans les pays occidentaux et dans des pays non-occidentaux. On peut donc considérer que l'idée de mérite est une idée morale hautement intuitive. Est-ce que cela constitue une justification suffisante de cette idée ?

À première vue, on pourrait être tenté de répondre par la négative. Les idées racistes sont très largement répandues, et pour autant il semble hautement contre-intuitif de considérer qu'il s'agit d'idées justifiées. Ou, plus précisément (parce qu'il n'est pas évident que le racisme soit une idée « morale ») : l'idée qu'il est moralement préférable de favoriser les membres de son groupe sur des membres d'autres groupes est une idée très largement répandue dans un très grand nombre de cultures. Pour autant, il ne semble pas évident que la préférence nationale soit par là justifiée. Est-ce que cela n'indique pas que les intuitions courantes n'apportent aucune aide pour justifier une théorie morale ?

Comme nous l'avons vu dans le chapitre 4, une telle conclusion serait prématurée. En effet, si l'on accepte une théorie méta-éthique non réaliste, on est obligé de conclure que toute théorie morale doit se fonder dans une certaine mesure sur les intuitions courantes. Les intuitions ordinaires constituent le fondement de toute théorie morale, et le rôle de la philosophie morale consiste à mettre en ordre nos différentes intuitions.

Dans ce cadre, on voit bien que le fait qu'une théorie soit intuitive ou largement répandue offre un premier type de validation de l'idée de mérite, à savoir une validation *prima facie*. Une validation *prima facie* consiste dans l'idée que, en l'absence de considérations opposées, une idée morale valide *prima facie* est temporairement considérée comme valable. Dans le cas de l'idée de préférence au sein du groupe, on pourrait faire l'hypothèse que, même si une telle idée pourrait posséder une certaine validité à première vue, il faut rejeter cette idée parce qu'elle entre en conflit avec d'autres intuitions fondamentales, comme l'égalité entre différents êtres humains. Dans le cadre de l'équilibre réflexif rawlsien, on peut soutenir que l'intuition de la préférence pour les membres de son groupe entre en conflit avec des principes moraux plus importants, et qu'il faut réviser cette intuition de bas niveau pour permettre la pleine application de principes de justice ou d'égalité¹⁷¹.

¹⁷¹ Bien sûr, il est également possible de défendre l'idée que la préférence pour les membres de son groupe est une idée morale valide, et devrait être conservée même dans le cadre de l'équilibre réflexif. Pour une version très modérée de cette thèse, on peut voir par exemple le récent livre de David Miller sur l'immigration, David MILLER, *Strangers in our midst: the political philosophy of immigration*, Harvard University Press, 2016.

Dans ce contexte, comment faire le tri entre les intuitions qu'il s'agit de conserver et celles dont il faut se débarrasser ? Comment faire la distinction entre des intuitions véritablement fondamentales, et celles qui doivent être ultimement rejetées ?

Un premier critère, que nous venons d'examiner, revient à se demander dans quelle mesure le mérite, ou tout autre idée morale, entre en conflit avec d'autres idées fondamentales. Même sans entrer plus avant dans les détails, je considère qu'il est assez improbable que le mérite entre fondamentalement en conflit avec d'autres idées morales. En effet, nous venons de voir que le mérite est en général justifié de façon instrumentale par le fait que récompenser le mérite conduit à une augmentation de la productivité, et donc potentiellement du bien-être général. Dans ce cadre, la récompense du mérite semble en parfait accord avec d'autres idées qui nous tiennent à cœur.

Cependant, même indépendamment de l'accord du mérite avec nos autres idées morales, je voudrais souligner que le mérite semble être une idée trop fondamentale pour qu'il soit possible de vouloir s'en débarrasser. Même s'il était possible de prouver que le mérite est en conflit réel avec l'idée d'égalité, il me semble que le caractère basique de ces deux idées devrait conduire à la recherche d'un pluralisme éthique, et non pas à la décision de l'élimination d'une de ces deux idées pour atteindre une plus grande cohérence morale.

Pour démontrer que l'idée de mérite n'est pas éliminable, je pense qu'il est important de considérer certains critères qui peuvent être mis en avant pour défendre qu'une intuition est fondamentale au point que toute théorie morale plausible doit être contrainte de la conserver. Avant même d'examiner le conflit entre différentes intuitions morales, il est possible d'examiner chaque intuition pour se demander dans quelle mesure elle possède un ancrage profond dans nos intuitions morales.

Cette recherche de critères permettant d'identifier des traits centraux de la moralité a une longue histoire. Avant de proposer des critères, inspirés par la réflexion du philosophe contemporain Shaun Nichols, je voudrais examiner la proposition d'une fondation de la théorie morale dans les intuitions défendues par Sidgwick dans ses *Méthodes de l'éthique*. La théorie de Sidgwick présente l'avantage de proposer des critères pour fonder la moralité à partir de trois axiomes moraux. L'examen, et le rejet, des critères proposés par Sidgwick nous permettra de montrer pourquoi une fondation axiomatique n'est pas souhaitable pour la justification d'intuitions morales, et pourquoi il est nécessaire de se contenter d'une justification plus faible d'intuitions fondamentales.

Les critères axiomatiques de Sidgwick dans les Méthodes de l'éthique.

Dans les *Méthodes de l'éthique*, Sidgwick considère que toute théorie morale doit fondamentalement reposer sur les intuitions individuelles. Une théorie considérant que tout système moral doit se fonder sur les intuitions peut être décrit comme un type d'intuitionnisme. Pour Sidgwick, cependant, tous les intuitionnismes ne se valent pas, et il va distinguer trois théories différentes, afin d'en rejeter deux des trois : « Nous pouvons distinguer l'intuitionnisme perceptuel, d'après lequel la connaissance intuitive porte toujours sur l'adéquation d'actions particulières. L'intuitionnisme dogmatique, pour lequel les règles générales du sens commun sont acceptées comme axiomatiques. L'intuitionnisme philosophique, qui tente de trouver une explication plus profonde pour les règles actuelles¹⁷². »

Reprenons chacun de ces trois cas, pour mieux discerner en quoi des intuitions peuvent jouer un rôle justificatif central dans chacune de ses théories. L'intuitionnisme perceptuel repose sur l'idée que nous avons une intuition directe de la justesse des cas particuliers, comme par exemple du fait qu'il est mal de ne pas rendre *Guerre et Paix* à votre grand-oncle qui vous l'a prêté il y a trois semaines. Sidgwick rejette l'idée qu'il existe véritablement des intuitions particulières qui ne se basent pas sur des règles plus générales (l'idée qu'il faut rendre des objets prêtés à leur propriétaire légitime, dans notre cas).

Une grande partie des *Méthodes de l'éthique* porte sur l'examen des deux autres types d'intuitionnisme, l'intuitionnisme dogmatique et l'intuitionnisme philosophique. L'intuitionnisme dogmatique défend que nous possédons une connaissance immédiate de vérités morales générales. Parmi ces idées morales, on pourrait ranger l'idée qu'il faut respecter l'égalité entre citoyens, ou qu'il est bon de récompenser les individus selon leur mérite, ou (si l'on considère le sens commun du XIXe siècle) qu'un État doit avant tout respecter les intérêts de ses propres concitoyens. Sidgwick propose de montrer que l'ensemble de ces règles sont incohérentes, et qu'il faut les dépasser pour atteindre des règles plus générales qui doivent être considérées comme absolument axiomatiques. Sidgwick propose ainsi que seules trois règles peuvent prétendre au titre d'axiomes moraux : la prudence, la bienveillance, et l'égalité purement formelle. La prudence correspond à l'idée que mon bien-être futur vaut autant que mon bien-être présent, l'égalité purement formelle à l'idée que des cas similaires doivent être traités de la même façon, la bienveillance à l'idée que le bien-être d'une autre personne vaut

¹⁷² H. SIDGWICK, *The Methods of Ethics*, op. cit., chap. VIII. Ce que Sidgwick appelle l'intuitionnisme dogmatique correspond à ce qu'on appelle l'intuitionnisme classique de Ross.

autant que le mien, et que je dois choisir l'action qui augmente le bien-être total. Sidgwick n'accepte ainsi aucun principe de justice parmi ses axiomes fondamentaux de la morale.

Dans ce cadre, Sidgwick propose quatre axiomes, qui doivent lui permettre de montrer que les idées du sens commun ne peuvent pas servir d'axiomes, mais doivent être interprétées dans un cadre plus général utilitariste. Les quatre axiomes identifiés par Sidgwick sont les suivants : « Nous exigeons d'un axiome qu'il soit (1) énoncé dans des termes clairs et précis, (2) réellement évident par lui-même, (3) qu'il n'entre pas en conflit avec d'autres vérités, et (4) qu'il soit soutenu par un consensus adéquat par des experts. Ces caractéristiques ne se trouvent pas parmi les maximes morales du sens commun¹⁷³. » En d'autres termes, Sidgwick considère que les morales du sens commun recèlent de nombreuses obscurités, qu'elles entrent en conflit les unes avec les autres, et par conséquent ne font l'objet d'aucun consensus. A l'inverse, les trois axiomes fondamentaux de prudence, de bienveillance, et d'égalité formelle sont clairs, évidents, n'entrent pas en conflit les uns avec les autres et sont soutenus par tous les experts. En conséquence, ils peuvent servir de fondation à une théorie morale, qui sera forcément, pour Sidgwick, utilitariste.

Un point préliminaire doit être noté. Tout d'abord, notre justification des intuitions liées au mérite prend place entre ce que Sidgwick nomme intuitionnisme dogmatique et intuitionnisme philosophique. Sidgwick ne semble pas envisager de juste milieu entre un conservatisme absolu vis-à-vis des règles actuelles, et une tentative de dépassement radical vers des règles plus générales. Dans notre cas, nous ne voyons pas pourquoi il y aurait à choisir entre adopter parfaitement les opinions populaires, ou les dépasser complètement. En d'autres mots, selon les termes de Vargas que nous avons repris dans le chapitre 4, nous nous gardons le droit de proposer une théorie qui soit intermédiaire entre une philosophie conservatrice et révisionniste du sens commun.

Mais le problème avec les critères proposés par Sidgwick est plus large. Il faut reconnaître qu'il s'agit de critères beaucoup trop exigeants, qui risquent de réduire la moralité au plus pur formalisme. Mon argument procédera en deux temps. Dans un premier temps, je montrerai pourquoi ces critères ne peuvent pas apparaître comme des critères plausibles pour distinguer les intuitions morales véritablement fondamentales. Dans un deuxième temps, je conclurai qu'il faut abandonner l'ambition axiomatique pour proposer une version plus modeste de la fondation intuitionniste de la morale.

¹⁷³ *Ibid.*, chap. XI.

Reprenons maintenant les quatre points de Sidgwick. Tous possèdent une certaine ambiguïté, qui rend impossible leur bonne application. Tout d'abord, l'idée qu'un axiome moral doit être énoncé dans des termes clairs et précis peut sembler évident. Cependant, il est délicat de dire dans quelle mesure un axiome possède un degré suffisant de précision et de clarté. Prenons l'exemple de l'idée de bienveillance proposée par Sidgwick, qui consiste selon lui en l'idée que le bien être d'une autre personne vaut autant que le mien, et que je dois choisir l'action qui augmente le bien être total. Dans ce cas, on voit bien que l'idée possède une certaine obscurité dans la marge d'applications qui lui est proposée. Comment comprendre l'idée de personne ? Comment comprendre l'idée de bien-être ? Un traitement véritable de ces concepts implique nécessairement une analyse étendue. Les termes peuvent sembler clairs, mais recèlent pour autant une marge d'interprétation considérable.

Ensuite, il est difficile de savoir en quoi consiste exactement l'idée que des principes doivent être évidents par eux-mêmes. Les principes énoncés par Sidgwick semblent relativement naturels, mais il est toujours possible de formuler des contre-objections. Si je pense qu'une autre personne possède un mérite inférieur au mien, alors il n'est plus du tout évident que je devrais accorder la même valeur à son bien-être. Sauf à accepter l'utilitarisme, il n'est en rien évident que je doive choisir l'action qui maximise le bien-être total.

Troisièmement, il n'est pas du tout clair qu'il soit préférable de choisir des axiomes de la moralité ne possédant aucun conflit entre eux. Cela serait le cas si l'on voulait bâtir une axiomatique mathématique, où l'idée de non-contradiction entre axiomes serait décisive. Cependant, dans le domaine moral, il n'est pas évident de savoir pourquoi il serait vital de construire une théorie dénuée de conflits entre principes fondamentaux. Comme l'a défendu avec force Bernard Williams dans ses articles, il est important de reconnaître que la moralité humaine peut receler des conflits insolubles, et une théorie qui ne reconnaîtrait pas la force des conflits moraux peut sembler parfaitement inhumaine¹⁷⁴. Refuser les conflits moraux peut bien conduire à une théorie aseptisée, comme l'utilitarisme ; mais il n'est pas du tout évident que cela constituerait un progrès moral, en tant que cela nous conduirait à l'élimination de la plupart de nos intuitions morales quotidiennes. Evidemment, c'est précisément la conclusion que veut nous faire tirer Sidgwick : étant donné que nos intuitions courantes entrent en conflit, elles doivent être dépassées dans une perspective plus englobante. Mais on pourrait tout aussi bien tirer la conclusion que, puisque nos intuitions entrent en conflit, il faut développer une perspective pluraliste, qui accorde une importance à chacune de nos intuitions fondamentales,

¹⁷⁴ Bernard Arthur Owen WILLIAMS, *Problems of the self: philosophical papers 1956 - 1972*, Cambridge, Cambridge Univ. Press, 1999.

tout en sachant qu'il ne sera jamais possible de construire une théorie morale qui soit complètement libérée de tout conflit possible.

Enfin, l'idée d'un consensus d'experts pose problème. Il se pose à la fois la question de la sélection des experts, et surtout, il n'est pas évident qu'une valeur contestée par certains experts ne puisse servir de fondement à une théorie morale. Il semble que toute valeur fondamentale peut faire l'objet d'une contestation ; les idées d'égalité ou de mérite ont certainement été critiquées par différents « experts ». Même si Sidgwick voudrait conclure que ce dissensus montre que ces idées ne peuvent être fondamentales, on pourrait tout aussi bien conclure qu'il est préférable de rejeter le critère d'un consensus d'experts pour déterminer quelles intuitions morales doivent être considérées comme primordiales.

Ces difficultés nous amènent à une impasse. Étant donné qu'il est difficile de considérer un seul critère moral qui satisfasse les critères de Sidgwick, il semblerait nécessaire d'admettre qu'il n'existe aucun axiome de la moralité. Cette conclusion peut sembler absurde. Il faut donc rejeter les critères posés par Sidgwick en tant que critères nécessaires pour identifier des axiomes moraux. Plus généralement, cependant, on peut tirer une leçon générale de cette discussion pour rejeter l'idée d'une axiomatique de la moralité.

Sidgwick cherche à construire une théorie morale déductive sur la base d'un petit nombre d'intuitions fondamentales, qui jouent le rôle d'axiomes. En mathématiques, un axiome constitue une hypothèse fondamentale sur laquelle se fonde l'ensemble des démonstrations ultérieures. Considérer des intuitions morales comme des axiomes revient à affirmer deux choses : (1) il existe des idées morales qu'il est absolument impossible de remettre en cause, et (2) il est possible de construire une théorie morale complète sur la base d'un petit nombre d'énoncés.

S'il était possible de construire une théorie morale axiomatique convaincante, je pense que les quatre conditions édictées par Sidgwick seraient certainement nécessaires. Cependant, il semble absurde d'exiger de toute théorie morale qu'elle prenne une forme axiomatique. Pour la plupart des questions morales, il est vain d'imaginer que des énoncés moraux puissent approcher la clarté des énoncés mathématiques, qu'il soit possible d'éviter de la même manière toute possibilité de conflit, et qu'il soit possible de trouver le même consensus régnant entre les experts. Il faut adopter une théorie des intuitions morales qui ne requière pas que celles-ci jouent le rôle d'axiomes.

La plupart des méthodes en philosophie morale ne requièrent pas que certaines idées possèdent une valeur axiomatique. Selon la méthode de l'équilibre réflexif de Rawls, il suffit

de trouver un équilibre entre intuitions précises et principes de haut niveau, sans qu'une seule intuition, ou un seul principe de haut niveau, possède une valeur absolue et irréfutable.

Pour développer le processus à l'œuvre dans un exercice philosophique tel que l'équilibre réflexif, il peut être utile de passer par un parallèle avec la théorie de la justification épistémique développée par Susan Haack dans son livre *Evidence and Inquiry*¹⁷⁵. La théorie de la justification épistémique développée par Haack nous servira de modèle pour penser la justification en théorie morale telle qu'elle est mise en œuvre dans un processus d'équilibre réflexif.

Dans *Evidence and Inquiry*, Haack tente de dépasser l'opposition entre deux théories rivales de la justification épistémique, le cohérentisme et le fondationnalisme. Le cohérentisme considère que des croyances sont justifiées s'il n'existe aucune contradiction entre elles. Dans sa forme pure, le cohérentisme n'attache donc aucune importance fondamentale à certaines croyances particulières. Une croyance est uniquement justifiée par les relations qu'elle entretient avec nos autres croyances.

Le fondationnalisme, à l'inverse, considère qu'une croyance est justifiée si elle peut être déduite de croyances considérées comme fondamentales. Le fondationnalisme est donc proche d'un type de raisonnement axiomatique, même si les croyances considérées comme fondamentales peuvent aussi bien consister dans des véritables axiomes mathématiques que dans des croyances empiriques (par exemple, comme les évidences tirées de la sensation). Les croyances fondamentales sont justifiées tout à fait indépendamment de leur relation avec d'autres croyances.

En opposition à ces deux théories de la justification, Haack propose une théorie alternative, qu'elle appelle le fondhérentisme (« foundherentism »). Voici comment Haack justifie la possibilité de construire une alternative à ces deux théories : « Exprimé de façon la plus simple possible, l'argument est le suivant : le fondationnalisme exige que la justification aille dans une seule direction ; le cohérentisme ne l'exige pas. Le cohérentisme exige que la justification soit exclusivement une question de relations entre croyances ; le fondationnalisme ne l'exige pas. [...] Donc : une théorie qui accorde une place à des données qui ne sont pas des croyances ne peut pas être cohérentiste ; une théorie qui n'exige pas que la justification soit unidirectionnelle ne peut pas être fondationnaliste. Une théorie comme celle que je défends, qui accorde de l'importance à l'expérience dans la justification, mais qui n'admet pas de classe de croyances privilégiées justifiées exclusivement par l'expérience sans aucun soutien procuré par d'autres

¹⁷⁵ Susan HAACK, *Evidence and inquiry: a pragmatist reconstruction of epistemology*, Prometheus Books, 2009.

croyances, n'est ni fondationnaliste ni cohérentiste, mais est intermédiaire entre les rivaux traditionnels¹⁷⁶. »

Haack caractérise le fondhérentisme, sa position intermédiaire, ainsi : « (FH1) L'expérience d'un sujet est pertinente pour la justification de ses croyances empiriques, mais il n'y a pas nécessairement de classe de croyances empiriques justifiées exclusivement par le soutien de l'expérience, indépendamment du soutien d'autres croyances ; (FH2) La justification n'est pas exclusive ou ne suit pas une seule direction, mais implique des relations systématiques de soutien mutuel¹⁷⁷. »

En d'autres termes, la position défendue par Susan Haack se base bien sur l'idée qu'il existe certaines croyances qui sont davantage fondamentales que d'autres, mais ne considère pas que ces croyances fondamentales aient un statut axiomatique, ou qu'il soit impossible de les remettre en cause. Pour Haack, même si certaines données ont un statut épistémique particulièrement important (notamment du fait qu'il s'agit de données directement issues de l'expérience), cela ne veut pas dire qu'il n'est pas possible de les remettre en cause si elles sont en conflit avec un trop grand nombre de données empiriques.

Pour comprendre l'importance d'adopter une position fondhérentiste, il faut reprendre les arguments les plus importants portés contre les positions fondationnalistes et cohérentistes. L'argument le plus fort contre la théorie cohérentiste porte sur le fait qu'un ensemble de croyances cohérentes peut être basé sur des fondations purement imaginaires. Il est possible d'avoir une croyance parfaitement cohérente dans l'existence d'un monde des fées, par exemple. Sans ajout de données empiriques, on peut se noyer dans le plus pur imaginaire, et il est douteux que des croyances puissent être dites justifiées dans ce cas. Dans le cas de la moralité, il est de la même manière possible de construire un système moral cohérent, sans que ce système moral soit particulièrement attirant, ou sans qu'il soit possible de le considérer comme un système moral plausible. Les théories utilitaristes sont parfaitement cohérentes ; elles sont généralement rejetées parce qu'elles entrent en contradiction avec nos intuitions ordinaires. Si l'on veut aller encore plus loin, on peut imaginer des systèmes moraux complètement absurdes et quand même cohérents. Si l'on imagine une espèce extra-terrestre qui considère que le but suprême de la moralité est de maximiser le nombre d'hexagones dans l'univers, alors on peut reconnaître qu'il peut s'agir d'un système moral cohérent, sans qu'il s'agisse d'un système moral vraisemblable, parce qu'il rompt avec nos intuitions morales ancrées.

¹⁷⁶ *Ibid.*, p. 19.

¹⁷⁷ *Ibid.*

En ce qui concerne la justification épistémique comme la justification morale, un système purement cohérentiste doit donc être rejeté. De la même manière, le fondationnalisme n'est pas acceptable parce qu'il postule l'idée de données parfaitement irréfutables, qu'il s'agisse de données empiriques ou mathématiques. Cependant, même dans le cas de croyances fondationnelles, l'existence d'autres arguments ou d'autres croyances doit pouvoir procurer un plus grand degré de justification à ces croyances. Reprenons l'exemple de Haack selon lequel une croyance fondationnelle plausible pourrait être le fait que je crois qu'un chien est en face de moi, parce que je le vois. Même dans ce cas, cette croyance serait encore davantage justifiée si je sais que mes yeux fonctionnent bien, si je sais que je ne suis pas en train de participer à une séance d'hypnose, si je sais qu'il n'y a pas de jouet ou de robot en forme de chien dans la maison, etc. Il n'existe pas de croyances si fondamentales qu'elles ne puissent pas être affaiblies ou renforcées par l'existence d'autres croyances. De la même manière, dans le domaine moral, on peut considérer que nos intuitions morales fondamentales sont renforcées si elles forment un tout cohérent avec nos autres intuitions et nos autres croyances éthiques.

En somme, de la même manière que Haack défend une perspective épistémique mélangeant des aspects cohérentistes et fondationnalistes, je pense qu'il est important de mêler des aspects de ces deux traditions dans une théorie de la justification morale. Dans la perspective épistémique défendue par Haack, une croyance sera d'autant plus justifiée qu'elle sera issue directement d'une donnée empirique, et d'autant plus qu'elle sera en cohérence avec un plus grand nombre d'autres croyances. Il existe une double source de la justification, à la fois due à la proximité avec des données empiriques et par l'harmonie entre cette croyance et les autres.

De la même manière, on peut considérer qu'il existe une double source de la justification d'une croyance morale. Une croyance morale est d'autant plus justifiée qu'elle est davantage proche d'intuitions fondamentales, et d'autant plus qu'elle est en plus grande harmonie avec d'autres croyances morales. Toute la question est alors de savoir comment identifier les croyances fondamentales.

La perspective ouverte par le fondhérentisme de Haack nous permet de voir qu'il n'est pas nécessaire, pour identifier des croyances fondamentales, de découvrir des idées morales qui soient parfaitement axiomatiques, et qui satisfassent l'ensemble des conditions proposées par Sidgwick. Une idée morale n'a pas besoin d'être parfaitement évidente par elle-même, et de faire l'objet d'un consensus parfait, pour être considérée comme étant une croyance morale importante. La perspective ouverte par Haack nous offre la possibilité de tenir deux choses ensemble : l'idée que certaines intuitions morales possèdent un très haut degré de justifications, et ont très peu de chances d'être remises en cause ; et l'idée qu'aucune idée n'est parfaitement

justifiée et axiomatique, c'est-à-dire que toute idée peut toujours être remise en cause si elle entre en conflit avec trop d'idées morales différentes. Dans un domaine épistémique, l'idée qu'un chien est devant moi semble très fortement justifiée si je possède tous les *sense data* nécessaires, mais peut quand même être révoquée en doute si je possède des croyances contradictoires (comme le fait que je viens de participer à une séance d'hypnose).

Des critères intéressants visant à identifier des intuitions morales fondamentales ont été apportés par Shaun Nichols dans son article « Brute Retributivism¹⁷⁸ ». Nichols ne considère pas que toutes les intuitions morales peuvent être considérées comme fondamentales. Il considère cependant qu'un sous-ensemble particulier d'intuitions peut être considéré comme possédant un statut privilégié, à savoir le sous-ensemble des normes profondément ancrées dans notre psychologie morale. Selon Nichols, « Les normes profondément ancrées ont trois traits. Premièrement, elles sont très largement répandues dans la communauté. Deuxièmement, des normes ancrées sont fondamentales d'un point de vue inférentielles, c'est-à-dire qu'elles ne sont pas dépendantes d'autres normes ou d'autres faits [...]. Enfin, des normes ancrées sont enracinées dans des émotions humaines ; elles entrent en résonance avec notre bagage émotionnel naturel¹⁷⁹. »

Nichols ne propose pas de justification générale de ces trois critères. Cependant, il est facile de justifier pourquoi elles possèdent une validité d'un point de vue normatif. L'idée que des normes fondamentales doivent être largement répandues au sein de la communauté semble un point de départ évident. Il serait difficile d'imaginer construire une moralité commune si les intuitions en question étaient tout à fait minoritaires.

On retrouve un critère proche de celui défendu par Sidgwick, quand il défendait qu'un axiome moral devait faire l'objet d'un consensus au sein d'un groupe d'expert. Le critère proposé par Nichols se distingue de deux manières de celui proposé par Sidgwick. Tout d'abord, Nichols ne fait plus appel à des experts, mais simplement au plus grand nombre. Cet élargissement nous semble un progrès, en tant qu'il est difficile d'imaginer une manière parfaitement valable de constituer un groupe d'experts en termes de moralité. La deuxième différence entre la version proposée par Nichols et celle de Sidgwick tient au fait que Nichols n'exige plus l'existence d'un consensus, mais requiert simplement que l'intuition soit très largement partagée. Là encore, il nous semble s'agir d'un progrès. Un consensus parfait semble très difficilement atteignable, puisqu'il existe toujours des opinions dissidentes. Cependant, il

¹⁷⁸ Nichols SHAUN, « Brute Retributivism », in Thomas A. NADELHOFFER (dir.), *The Future of Punishment*, Oxford University Press, 2013, p. 25-46.

¹⁷⁹ *Ibid.*

est vrai qu'il existe des opinions très largement partagées. Sans chercher à faire de dichotomie, on peut alors considérer qu'une opinion sera d'autant plus ancrée qu'elle sera partagée par un plus grand nombre de personnes différentes.

On peut également chercher à redoubler le critère de popularité proposé par Nichols. Nichols cherche à identifier des critères ancrés au sein d'une communauté particulière. Mais on peut également considérer qu'un autre indice de l'ancrage d'une intuition morale peut être la popularité d'une intuition dans un très grand nombre de communautés différentes. Si une idée morale n'était populaire qu'au sein d'une seule communauté humaine, alors on pourrait considérer qu'il s'agirait d'un bien piètre candidat au titre d'intuition fondamentale, en tant que des changements sociétaux ou culturels auraient de fortes chances de renverser cette idée morale. A l'inverse, si une idée morale se retrouvait à l'identique dans un très grand nombre de cultures, alors il s'agirait d'un bon indice du fait que l'idée est robuste et ne peut être changée facilement.

Le lecteur aura compris où je veux en venir : le mérite remplit parfaitement ces deux premiers critères de validité. Comme nous l'avons vu dans le chapitre 3, le mérite fait l'objet d'un très fort soutien dans un très grand nombre de pays. Il remplit les critères de popularité aussi bien au sein des pays occidentaux, qu'au sein d'un certain nombre de pays non-occidentaux. En conséquence, il est très improbable que cette intuition si profondément ancrée soit renversée à l'issue d'un exercice d'équilibre réflexif.

Le deuxième critère avancé par Nichols correspond à l'idée qu'une intuition ancrée doit être fondamentale d'un point de vue inférentiel. Ce critère se comprend facilement : il s'agit de l'idée qu'une norme ne doit pas intervenir uniquement comme moyen instrumental de promouvoir une autre idée morale. Par exemple, l'idée qu'il est moralement important de trier nos déchets n'est pas fondamentale d'un point de vue inférentiel : nous justifions cette norme à partir de l'importance morale de protéger la planète, ce qui peut à son tour être potentiellement justifié par l'importance de protéger le bien-être des êtres vivants.

Dans le cas du mérite, nous avons argumenté dans la première partie de ce chapitre en faveur de l'idée que les intuitions ordinaires sont incompréhensibles si l'on fait l'hypothèse que le mérite possède une valeur purement instrumentale. Si l'on pensait que la récompense du mérite est seulement justifiée parce qu'elle sert la promotion du bien-être collectif, par exemple, alors l'on ne comprendrait pas pourquoi il serait bon d'un point de vue moral que des individus désintéressés bénéficient de leurs bonnes actions. Étant donné que le mérite ne se réduit pas à un pur rôle instrumental, il semble fondamental d'un point de vue inférentiel, et satisfait donc le deuxième critère de Nichols.

Le troisième critère invoqué par Nichols, enfin, est que les intuitions fondamentales sont « enracinées dans des émotions humaines ; elles entrent en résonance avec notre bagage émotionnel naturel ». Un tel critère n'est pas en soi incontestable, étant donné qu'il serait possible d'imaginer qu'une théorie morale rationaliste n'accorde aucune importance à l'idée que des émotions particulières soient liées à une intuition morale. Il s'agit bien sûr d'un critère en phase avec la propre théorie de Nichols, qui considère que les émotions constituent une part essentielle de tout jugement moral, en concordance avec des jugements de normes¹⁸⁰. Sans entrer dans des débats psychologiques complexes sur le rôle de la raison et des émotions dans la production de jugements moraux, il est assez facile de défendre l'utilité, au moins heuristique, du critère proposé par Nichols. Quoique l'on pense du rôle des émotions dans la psychologie morale, il semble évident que les émotions jouent un rôle fort dans la motivation à l'action. Il semble également assez net que le contrôle individuel sur les émotions n'est pas infini, et qu'il peut être difficile de changer fondamentalement les émotions suscitées spontanément.

Si l'on reconnaît le rôle motivationnel des émotions, leur impact sur l'importance que nous accordons à chaque idée morale, leur malléabilité limitée, il faut conclure qu'il sera généralement plus difficile d'éliminer une intuition à laquelle sont liées des émotions fortes qu'une intuition à laquelle ne sont pas liées des émotions puissantes. Ce critère de difficulté n'est pas en soi suffisant ; s'il y avait des raisons morales fortes pour critiquer une idée morale émotionnellement chargée, alors il devrait pouvoir être possible de le faire (à l'inverse, même si des animaux laids peuvent n'apporter aucun sentiment d'empathie, on peut considérer qu'il est fondamental de les préserver et d'empêcher leur souffrance). Cependant, cela ajoute de la plausibilité à l'idée qu'une idée morale soit fondamentale si elle est liée à des intuitions puissantes.

Maintenant que l'importance des émotions a été établie pour déterminer si une intuition est fondamentale, dans quelle mesure peut-on dire que les jugements de mérite suscitent de fortes émotions ? Le chapitre 3 s'est concentré sur les jugements cognitifs, et j'ai très peu insisté sur le rôle des émotions suscitées par la récompense des jugements de mérite, ou plutôt, sur les possibles émotions négatives suscitées par sa non-récompense. Cependant, la littérature psychologique et sociologique ne laisse aucun doute sur les sentiments d'indignation qui

¹⁸⁰ Shaun NICHOLS, « Norms with feeling: towards a psychological account of moral judgment », *Cognition*, 2002, vol. 84, n° 2, p. 221-236.

peuvent être provoqués par l'absence de récompense de bonnes actions, ou par l'absence de punition de crimes¹⁸¹.

Dans cette section, nous avons présenté un ensemble de critères qui indiquent que le mérite peut compter comme une intuition morale fondamentale. Il s'agit d'une idée qui fait l'objet d'un consensus fort, dans un très grand nombre de pays. Il s'agit d'une idée qui est fondamentale d'un point de vue inférentiel, c'est-à-dire qui n'est pas dérivée d'autres principes de justice. Enfin, il s'agit d'une idée dont la non-satisfaction provoque des émotions fortes, notamment dans le domaine de l'indignation. Pour toutes ces raisons, il est extrêmement douteux que l'on puisse s'en débarrasser dans le cadre d'un processus d'équilibre réflexif.

Nous avons également rejeté dans cette section une conception axiomatique de la morale, comme celle proposée par Henry Sidgwick. Nous ne considérons donc pas qu'il existe des idées morales qui soient si évidentes qu'il serait impossible de s'en débarrasser. Cependant, en l'absence de raisons fortes et convaincantes à l'encontre de l'idée de mérite, cette dernière semble posséder une validité à première vue. Le fait que cette intuition soit tellement ancrée lui offre un certain type de justification morale. Et, comme nous l'avons vu et allons le voir, il existe plusieurs autres justifications de l'idée de mérite, qui font que nous n'avons fondamentalement aucune raison d'aller à l'encontre de nos intuitions ancrées. Abordons maintenant le troisième type de justification en faveur de l'idée de mérite : le fait que la récompense du mérite assure un ordre entre l'intérêt personnel et l'intérêt collectif.

Le mérite comme mise en cohérence de l'intérêt personnel et collectif

La troisième justification tient au fait que la récompense du mérite assure une concordance, ou adéquation, entre intérêt personnel et intérêt public. La récompense selon le mérite garantit un ordre des valeurs, qui assure à l'individu l'existence d'une concordance entre ses raisons intéressées et ses raisons désintéressées.

Si, quand un individu effectue une action qui conduit au bien général, cet individu est récompensé, alors il existe une identité d'intérêt entre l'acteur et les autres membres de la

¹⁸¹ Francois DUBET, *Injustices. L'expérience des inégalités au travail*, Paris, Points, 2013 ; Daniel KAHNEMAN, David SCHKADE et Cass SUNSTEIN, « Shared Outrage and Erratic Awards: The Psychology of Punitive Damages », *Journal of Risk and Uncertainty*, 1998, vol. 16, n° 1, p. 49-86 ; Daniel KAHNEMAN et Cass SUNSTEIN, « Indignation: Psychology, Politics, Law », *Law & Economics Working Papers*, 2007.

société. A l'inverse, si un individu effectue une action généreuse et paye un coût pour cela, alors il existera un choix à effectuer entre la préservation de l'intérêt individuel et le bien-être collectif. La récompense du mérite assure qu'il existe un maximum de cas où l'intérêt individuel est en harmonie avec l'intérêt collectif.

Pour mieux comprendre quel est le rôle de la récompense du mérite, et en quoi il s'agit d'un cas d'harmonisation entre intérêt privé et collectif, il peut être utile de reprendre la matrice des interactions sociales possibles, très couramment utilisée en théorie des jeux et en biologie évolutionniste (tableau 1)¹⁸². Un comportement qui est bénéfique à la fois pour l'acteur et le récepteur correspond à un cas de bénéfice mutuel. On peut penser à une situation de coopération où les deux individus gagnent au change. Un comportement bénéfique pour l'acteur, mais négatif pour le récepteur, correspond à une action égoïste. On peut penser à un cas de vol. Un comportement bénéfique pour le récepteur, mais négatif pour l'acteur, est un comportement altruiste. On peut penser au fait d'aider quelqu'un, sans être récompensé en retour. Un comportement qui est négatif à la fois pour l'acteur et pour le récepteur est un cas de malveillance. On peut penser à un cas de sabotage d'un projet collectif, alors même que la personne qui a choisi de saboter le projet perd au fait d'avoir effectué le sabotage.

Tableau 1. Classification des comportements sociaux en fonction de leur impact sur l'acteur et le récepteur.

	Effet sur le récepteur	
	+	-
Effet sur l'acteur		
+	Bénéfice mutuel	Egoïsme
-	Altruisme	Malveillance

A partir de ce cadre conceptuel, on peut comprendre ce que cela fait de récompenser les actions favorables à la collectivité. Une action pro-sociale non-récompensée est une action altruiste, négative pour l'acteur et positive pour la collectivité. Une action pro-sociale

¹⁸² Ce tableau est tiré de la classification des interactions sociales présentée dans S. A. WEST, A. S. GRIFFIN et A. GARDNER, « Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection », *Journal of Evolutionary Biology*, 2007, vol. 20, n° 2, p. 415-432, p. 418. L'idée d'une telle classification est très courante, mais le vocabulaire employé peut varier d'un auteur à l'autre.

récompensée est une action qui relève du bénéfice mutuel, où à la fois l'acteur et la communauté bénéficient des actions qui ont été effectuées.

On peut comprendre de la même manière l'impact que la punition a sur les incitations individuelles. Si un individu fait une action égoïste, comme le vol, alors il récolte un bénéfice, au prix d'un impact négatif sur d'autres individus. La punition a alors pour fonction de s'assurer d'une adéquation entre les incitations individuelles et collectives : l'action du vol devient négative pour tout le monde, y compris pour la personne qui a commis le vol. En général, la récompense (ou la punition) selon le mérite conduit à passer d'une situation de conflit entre l'intérêt personnel et l'intérêt collectif à une situation d'accord entre les deux.

Bien sûr, la récompense du mérite ne conduit pas toujours à passer d'une catégorie de conflit d'intérêts à une catégorie d'harmonie d'intérêts. Par exemple, si un individu effectue une action qui est favorable à l'intérêt collectif, mais qui sert aussi son intérêt individuel, on peut considérer qu'il mérite de gagner davantage. Par exemple, un grand écrivain peut très bien avoir gagné beaucoup d'argent grâce à ses droits d'auteur, et quand même mériter un prix Nobel de littérature. Dans ce cas, on part d'une situation qui est de bénéfice mutuel, et on arrive (avec l'attribution du Nobel) à une situation qui reste dans la même catégorie, celle de bénéfice mutuel. De la même manière, à l'autre extrême, un acte de malveillance, comme un acte de sabotage, peut très bien mériter une forte punition, et ce, même si l'agent n'avait rien gagné à avoir effectué un acte de sabotage. Ceci étant dit, l'impact général d'une récompense conforme au mérite est bien d'aligner l'intérêt personnel et l'intérêt collectif, qui pourraient être en conflit. Une fois que l'on a dit cela, en quoi cela constitue-t-il une nouvelle justification du mérite ?

Tout d'abord, il faut remarquer qu'il ne s'agit pas entièrement d'une nouvelle forme de justification du mérite. Et ce, pour deux raisons. Tout d'abord, l'idée d'un alignement entre intérêt individuel et intérêt collectif est simplement la traduction d'un point de vue individuel du fait que la récompense du mérite a généralement un effet incitatif. Ensuite, le fait que nous avons l'intuition que c'est un mal s'il n'y a pas d'adéquation entre intérêts privés et intérêts collectifs est, à nouveau, la traduction du fait que nous avons des intuitions fortes que la récompense du mérite est une bonne chose. En d'autres termes, l'idée que le mérite est justifié par le fait qu'il instaure un ordre entre intérêt personnel et collectif correspond en grande part à une reformulation des deux premières justifications du mérite. Cependant, cette reformulation offre la possibilité d'un nouveau regard, et à ce titre constitue véritablement une nouvelle justification du mérite. Il est en effet frappant de noter à quel point l'idée d'un désaccord entre intérêt personnel et intérêt collectif constitue une idée centrale dans l'histoire de la philosophie morale et dans certaines œuvres littéraires.

La recherche d'une harmonie entre intérêt privé et intérêt public se retrouve de façon constante dans l'histoire de la philosophie morale. Il s'agissait d'une préoccupation centrale de la philosophie grecque, notamment à travers la recherche des raisons pour lesquelles la justice constituait un bien pour l'individu dans la *République* de Platon. Plus récemment, la question de la congruence entre sens de la justice et bien individuel a constitué un thème central de la troisième partie de la *Théorie de la justice*¹⁸³. Comme Christine Korsgaard le souligne dans les *Sources de la normativité*, toute la philosophie morale britannique du XVIIIe siècle est consacrée à la question de l'harmonie possible entre intérêt personnel et contributions morales¹⁸⁴.

Bien sûr, le fait que de nombreux philosophes ont considéré qu'il était important de résoudre les possibles conflits entre intérêt personnel et intérêt collectif ne veut pas dire que cette résolution des conflits par le biais de la récompense du mérite constitue une justification valable du mérite. De plus, la question de l'adéquation entre intérêt personnel et intérêt collectif, telle qu'elle a été étudiée par de nombreux philosophes classiques, n'était pas forcément la question d'un possible lien empirique entre bonnes actions et récompense, mais davantage la question de la recherche d'un lien intrinsèque. En d'autres termes, les philosophes classiques se posaient la question d'un lien *nécessaire* entre actions positives pour le groupe et intérêt personnel. Evidemment, la réponse correcte est qu'il n'existe aucune garantie d'une nécessaire harmonie, comme Sidgwick l'a montré de façon magistrale il y a plus d'un siècle dans *The Methods of Ethics*. Cependant, dans les limites pratiques, la récompense du mérite peut servir à assurer qu'il existe une cohérence maximale entre intérêt personnel et collectif, et, dans ce cadre, il s'agit d'une réponse pertinente aux inquiétudes philosophiques classiques.

De façon plus générale, il est possible de voir dans l'impératif d'harmonisation entre intérêts personnels et intérêts publics la trace d'une question philosophique plus large, à savoir la nécessité d'une harmonisation entre différents ordres de valeur. Selon toute conception plausible du bien de l'individu, chaque personne possède une pluralité d'intérêts et une pluralité d'identités, qui sont la source d'une pluralité d'obligations. En tant que citoyens, nous possédons des obligations politiques, comme par exemple voter et payer nos impôts ; en tant qu'êtres humains, nous possédons des devoirs minimaux de décence et d'assistance mutuelle en cas de danger ; en tant que membres d'une famille et de réseaux d'amitiés, nous possédons

¹⁸³ Le § 86 de la *Théorie de la justice* est tout entier consacré à la question de la congruence entre l'intérêt individuel et la recherche de la justice collective.

¹⁸⁴ C. M. KORSGAARD, *The Sources of Normativity*, *op. cit.* Cette recherche d'une adéquation entre individu et collectif est très nette par exemple chez Hume.

des devoirs spéciaux vis-à-vis de nos proches. De plus, même si l'on ne parle pas souvent de « devoirs » dans ce cas, il semble évident que toute vie bonne doit passer par un minimum d'attention à soi, que ce soit dans une culture personnelle ou dans la recherche du bonheur.

Toute vie bonne passe donc par le besoin de satisfaire une multitude d'obligations. Et, dans ce cadre, il est possible qu'il existe un ordre dans ces obligations, de telle sorte que nos différents besoins vitaux soient en harmonie les uns avec les autres. Mais il est également possible qu'il existe des conflits, et une impossibilité de satisfaire certains de nos devoirs sans en violer d'autres. Une telle situation semble constituer le summum de l'injustice, comme dans l'histoire classique de l'individu qui a le choix entre voler des médicaments, ou voir sa femme mourir de maladie.

La formulation la plus éloquente de l'importance de l'ordre au sein de nos différentes valeurs se trouve au début de *l'Enracinement*. Simone Weil y écrit : « Le premier besoin de l'âme, celui qui est le plus proche de sa destinée éternelle, c'est l'ordre, c'est-à-dire un tissu de relations sociales tel que nul ne soit contraint de violer des obligations rigoureuses pour exécuter d'autres obligations. L'âme ne souffre une violence spirituelle de la part des circonstances que dans ce cas. Car celui qui est seulement arrêté dans l'exécution d'une obligation par la menace de la mort ou de la souffrance peut passer outre, et ne sera blessé que dans son corps. Mais celui pour qui les circonstances rendent en fait incompatibles les actes ordonnés par plusieurs obligations strictes, celui-là, sans qu'il puisse s'en défendre, est blessé dans son amour du bien¹⁸⁵. » Weil décrit de façon extrêmement forte l'idée que nous avons un besoin d'ordre, un besoin *spirituel* qui correspond à l'idée que toute violation de nos obligations représente une sorte de souillure morale.

Dans ce passage, Simone Weil ne considère bien sûr pas que le sacrifice de l'intérêt personnel puisse représenter une rupture si nette de nos besoins spirituels. On peut cependant considérer qu'une société dictatoriale où l'individu n'a que le choix entre la « menace de la mort » et l'aspiration à faire respecter des règles de justice correspond à une violation nette des besoins individuels d'ordre. De façon moins dramatique, une société où il n'y aurait aucune récompense du mérite des individus verrait les personnes choisir en permanence entre leurs besoins personnels et ceux du public et des autres individus. Un tel choix peut apparaître autrement plus tragique si l'on prend en compte que les intérêts personnels n'incluent pas forcément seulement des intérêts égoïstes, mais aussi ceux les intérêts des individus proches comme les membres de la famille.

¹⁸⁵ Simone WEIL, *Œuvres*, Paris, Gallimard, 2000, p. 1031.

L'idée d'ordre entre les différentes exigences morales qui s'imposent aux individus ne possède pas du sens dans toutes les théories morales. Si l'on adopte une théorie utilitariste purement impartiale, comme celle défendue par exemple par des utilitaristes classiques comme Bentham ou Sidgwick, qui défend l'idée que la bonne action est celle qui maximise le plaisir collectif, alors il n'existe pas de conflit entre deux obligations opposées. L'obligation est toujours une : faire l'action qui maximise le bonheur collectif. Pour concevoir que l'idée d'ordre entre nos obligations morales ait un sens, il faut accepter que nous disposions de plusieurs obligations morales différentes ; en un mot, il faut accepter un pluralisme de valeurs. Dans un article classique, Bernard Williams avait mis en avant le fait que toute théorie morale plausible doit accepter que les obligations morales ne disparaissent pas simplement parce qu'elles sont en conflit¹⁸⁶. Dans le cas de croyances en conflit, il est obligatoire d'abandonner une des deux croyances, s'il y a réellement une impossibilité entre les deux. Dans le cas des obligations, celles-ci ne disparaissent pas simplement parce qu'elles entrent en conflit les unes avec les autres. Si j'ai le choix entre voler mon prochain et nourrir ma famille, alors il s'agit d'un choix tragique, où il est impossible de rester parfaitement pur. L'obligation vis-à-vis de mes proches n'annule pas purement et simplement mes obligations vis-à-vis de mes concitoyens. Les théories utilitaristes, qui nient la possibilité d'un conflit de valeur, sont insatisfaisantes en ce qu'elles nient l'évidence d'un conflit entre nos différentes obligations. En ce sens, une théorie morale doit pouvoir accorder de l'importance à l'ordre, et à une possible résolution des conflits entre différents devoirs. Dans le cas du mérite appliqué au conflit entre le vol et la famille affamée, l'idée d'ordre correspond à l'idée d'une récompense du fait de respecter la loi et d'être un bon citoyen.

L'idée d'ordre est tellement centrale dans le cadre de notre cognition morale qu'elle a fait l'objet d'un certain nombre d'œuvres littéraires. La question du conflit entre nos obligations morales constitue notamment le cœur d'un grand nombre des pièces de Bertold Brecht, et visent à illustrer les incohérences morales auxquelles sont condamnés les individus dans des sociétés capitalistes. Dans *la Vie de Galilée*, Galilée doit faire le choix entre la poursuite de la vérité et le confort personnel, et choisit en conséquence de trahir ses idéaux scientifiques et politiques. Dans *Sainte Jeanne des Abattoirs*, Jeanne est confrontée à un choix entre le maintien de la pureté de ses intentions et l'action politique qui passe nécessairement, dans le contexte de la

¹⁸⁶ Bernard Arthur Owen WILLIAMS, « Ethical consistency », in *Problems of the self: philosophical papers 1956 - 1972*, Cambridge, Cambridge Univ. Press, 1999, p. 166-186.

pièce, par des mains sales et le recours à la violence. Jeanne choisit de maintenir sa pureté personnelle, au prix d'une répression sanglante des velléités révolutionnaires¹⁸⁷.

Il existe une pièce de Brecht qui est toute entière consacrée au conflit entre intérêt personnel et décence morale, et qui vise à démontrer pourquoi le système capitaliste conduit à la misère les individus qui veulent maintenir un certain degré de décence et de bonté. Cette pièce, *La bonne âme du Se-Tchouan*¹⁸⁸, constitue un exemple littéraire très fort de représentation du tragique d'une société qui pénalise les individus pour leurs actions méritoires. L'héroïne de la pièce, Shen Té, est une personne qui souhaite maintenir une décence morale dans un contexte de misère extrême. Séduits par son honnêteté et sa bonté, les Dieux lui donnent un peu d'argent, afin de prouver qu'il est possible d'être à la fois bon et prospère en Chine. Malheureusement, les espoirs des Dieux sont déçus. Shen Té est réduite à la misère par sa bonté, parce qu'elle se fait exploiter dans ses relations commerciales et par sa famille. La famille de Shen Té est dans une misère authentique, mais, si Shen Té voulait vraiment les aider, elle se retrouverait à son tour dans une situation de pauvreté extrême. Elle se trouve prise dans une situation où trois réseaux d'obligations coexistent : les devoirs vis-à-vis de sa famille, les devoirs d'honnêteté et de décence vis-à-vis des inconnus, et les obligations de préservation d'elle-même. Prise dans ses contradictions, elle se fait passer pour son cousin, Shui Ta, afin que ce double assume les relations de dureté et d'exploitation nécessaires à la survie de toute entreprise dans un système capitaliste. La conclusion de la pièce est simple : le système capitaliste est une torture, parce qu'il force chacun à divorcer de ses idéaux moraux, et à survivre en jouant de son mieux le jeu de l'exploitation du prochain.

Même si la pièce de Brecht représente de façon la plus claire les tragédies qui peuvent découler d'une situation de non-récompense du mérite, il n'est pas besoin d'aller chercher des situations fictionnelles pour trouver l'impact du désordre et du manque d'harmonie entre intérêt personnel et intérêt collectif. Un cas extrêmement net est représenté par les politiques d'incitations de certaines grandes industries au temps des trente glorieuses. Dans un certain nombre d'usines, les ouvriers disposaient d'un système d'incitation aux pièces, qui faisait qu'ils étaient payés en fonction du nombre de pièces fabriquées chaque jour, avec un malus s'ils en fabriquaient moins que le quota du jour, et un bonus s'ils en fabriquaient plus.

Jusqu'ici, tout va bien. Un système d'incitation qui récompense davantage ceux qui font un meilleur travail est conforme à des principes de mérite, et semble relativement juste. Là où

¹⁸⁷ Bertolt BRECHT, *La vie de Galilée*, Paris, L'Arche, 1997 ; Bertolt BRECHT, *Sainte Jeanne des abattoirs*, Paris, L'Arche Editeur, 2008.

¹⁸⁸ Bertolt BRECHT, *La Bonne Ame du Se-Tchouan*, trad. fr. Marie-Paule RAMO, Paris, L'Arche éditeur, 2004.

commence l'injustice et l'absurdité des politiques industrielles, c'est ce qui se passait dans le cas où des ouvriers dépassaient durablement les quotas assignés. En effet, si un groupe dépassait régulièrement les quotas, obtenant ainsi un bonus régulier, les quotas étaient augmentés, forçant les individus à travailler beaucoup plus pour obtenir le même salaire que ce qu'ils recevaient auparavant¹⁸⁹. Dans ce cas, les ouvriers étaient en réalité punis (par une baisse du salaire à la pièce) parce qu'ils avaient contribué de façon positive à la réussite de l'entreprise (en dépassant de façon durable les quotas établis).

Une telle situation n'était qu'en apparence incitative. Le système mis en place par les entreprises incitait effectivement les ouvriers à atteindre leurs quotas, et punissait effectivement les ouvriers qui faisaient moins que le quota assigné. Cependant, ce système punissait également les ouvriers qui faisaient trop bien leur travail, fournissant ainsi une incitation très forte à ne pas trop travailler.

Si l'on réanalyse cette situation à partir de l'idée d'ordre ou de congruence entre les valeurs, on peut considérer que les ouvriers disposaient de trois types d'obligations ou d'intérêts spéciaux dans cette situation. Le premier intérêt était l'intérêt personnel, correspondant à l'idée qu'il était bon pour eux de tirer le meilleur salaire de cette situation. La deuxième obligation correspondait au fait qu'il était bon de fournir une activité productive et de participer à la réussite de l'entreprise. La troisième obligation était de ne pas mettre en danger le salaire des collègues, et de ne pas contribuer à l'exploitation ouvrière.

Ces trois intérêts fondamentaux étaient en opposition fondamentale. Sur le court terme, le système d'incitation assurait une congruence entre l'intérêt personnel et la contribution au bien-être de l'entreprise. Mais cette congruence n'était qu'à court terme : si trop d'ouvriers travaillaient trop bien, alors tous étaient punis. Sur le long terme, il y avait une opposition entre l'intérêt personnel, l'intérêt collectif des salariés, et le fait de contribuer à la réussite de l'entreprise. En conséquence, les ouvriers commençaient à arrêter de travailler, ou à travailler de façon moins intense, à partir du moment où ils avaient atteint leur quota. Les dernières heures de travail pouvaient devenir des moments de pause, ou d'absence de travail, pour éviter d'effectuer un travail de trop bonne qualité.

¹⁸⁹ Une telle situation est décrite de façon particulièrement forte dans les articles classiques de Donald ROY, *Un sociologue à l'usine*, Paris, La Découverte, 2006. Robert Linhart, dans *l'établi*, décrit la même situation. La situation est évoquée dans un épisode de la série de Fassbinder, *Huit heures ne font pas un jour*, où des ouvriers sont punis par une baisse de salaire pour avoir inventé une machine qui simplifiait leur travail. Robert LINHART, *L'établi*, Paris, Editions de Minuit, 1981 ; Rainer Werner FASSBINDER, *Huit heures ne font pas un jour*, Carlotta Films.

Une telle situation peut être analysée de nombreuses manières. Il est possible bien-sûr de critiquer ces systèmes d'incitation sur une base égalitariste. Si les ouvriers voient leur salaire baisser à chaque fois qu'ils dépassent un quota, alors il est impossible qu'ils améliorent leur salaire. La politique de ces entreprises condamnait concrètement les ouvriers à un salaire bas sur le long terme.

Bien sûr, la critique égalitaire est parfaitement pertinente. Mon argumentation dans ce chapitre suggère tout de même que ce n'est pas le seul problème moral à l'œuvre ici. Nous sommes révoltés par ce genre de politiques en partie parce qu'elles instaurent un conflit fondamental entre la recherche de bénéfices personnels et la contribution à l'efficacité collective. Dans ce cas, pour préserver leurs intérêts, les ouvriers devaient limiter leurs efforts et faire un travail bien moins productif que ce dont ils étaient capables. Ce désordre et ce désaccord entre l'intérêt personnel et collectif est fondamentalement lamentable. Un système parfaitement ordonné est un système qui assure que les individus ne seront pas punis pour leurs bonnes actions. En l'occurrence, dans le cas de l'entreprise, un système de rémunération conforme au mérite aurait correspondu à une situation où il n'y aurait pas eu de risque de baisse de salaire à la suite de l'amélioration de la qualité du travail.

Conclusion

Nous avons vu dans ce chapitre trois justifications principales du mérite. La première justification tient au fait que, de façon générale, la récompense du mérite crée un système d'incitation qui augmente la productivité des individus. Même si une telle justification apporte un soutien appréciable à l'idée de mérite, elle n'est pas en elle-même suffisante, parce qu'il est toujours possible de trouver des cas où récompenser des individus productifs n'aurait aucun impact sur l'efficacité générale. Une compréhension purement utilitaire du mérite dirait qu'il n'est dans ce cas pas nécessaire de récompenser ces individus, tandis qu'une conception qui lui accorde une valeur intrinsèque dirait qu'il est bon de récompenser des individus méritants même quand cette récompense ne sert aucun but ultérieur.

Une deuxième justification, intrinsèque, est donc nécessaire. Nous avons défendu l'idée que le mérite est une idée morale extrêmement ancrée dans l'homme, ce qui se voit par le fait que l'idée possède un degré très fort de consensus aussi bien au sein de chaque pays qu'à travers différents pays, par le fait qu'il s'agit d'une idée première et non dérivée, et par le fait qu'elle

suscite des émotions fortes. Dans ce cadre, et en l'absence de raisons opposées, il est possible de dire que ce qui semble universellement moralement bon doit être moralement bon.

Enfin, nous avons vu qu'une troisième justification du mérite est l'instauration d'une concordance entre intérêt personnel et intérêt collectif. Il s'agit, en d'autres termes, de la version individualisée de l'idée que la récompense du mérite correspond à la création d'un bon système d'incitations. Le mérite assure qu'il n'y a pas de désordre et de contradiction entre deux valeurs morales auxquelles nous tenons : la recherche de l'intérêt personnel et la poursuite du bien-être collectif.

Nous allons voir maintenant la quatrième justification possible de la récompense du mérite : récompenser le mérite constitue un système de communication qui transmet des informations aux individus sur la valeur de leur contribution.

Chapitre 7

Justifier le mérite par sa valeur communicative : Introduction aux débats théoriques

Dans ce chapitre, je procéderai à la présentation de la quatrième justification du mérite, basée sur l'idée que récompenser le mérite sert à communiquer l'approbation de la société à l'individu récompensé. Une théorie communicative du mérite est relativement peu explorée en ce qui concerne le mérite positif. Peu d'auteurs ont défendu que la communication pouvait servir à justifier l'attribution de récompenses. A l'inverse, en ce qui concerne le mérite négatif et la punition, cette voie a été plus qu'explorée au cours des cinquante dernières années, notamment depuis l'article fondamental de Feinberg « The Expressive Theory of Punishment » publié en 1965¹⁹⁰.

Il est facile de trouver des exemples de cas où la récompense au mérite joue un rôle communicatif, ou bien des cas où la punition exprime la réprobation sociale. Si un individu reçoit un bonus pour un travail bien fait, il interprétera sans doute cela comme un signe d'approbation de son travail. Si quelqu'un reçoit un salaire très faible, cette personne pensera sans doute que son travail n'est pas apprécié. De la même manière, la condamnation pénale exprime l'idée que le condamné a effectué une chose considérée comme inacceptable par la société. Le fait que les pratiques de récompense ou de punition puissent être interprétées selon un prisme communicationnel n'est pas en soi controversé. Ce qui est controversé, c'est de savoir dans quelle mesure cet aspect communicationnel peut servir de justification adéquate de la pratique de récompense ou de punition.

Dans ce chapitre, j'aborderai essentiellement la question de la justification du mérite à partir de la considération d'une théorie de la peine, contrairement aux précédents chapitres, où la plupart des exemples avaient été tirés de la pratique du mérite positif et de la récompense. Le principal apport d'un tel changement de perspective réside dans le fait que nous pourrions ainsi

¹⁹⁰ Joel FEINBERG, « The Expressive Function of Punishment », *The Monist*, 1965, vol. 49, n° 3, p. 397–423.

exploiter la littérature d'une extrême richesse sur le sujet. Les théories de la peine ont en effet abordé de multiples manières en quoi la punition pouvait avoir une valeur expressive. On peut trouver, de plus, de nombreuses objections de qualité à la théorie communicationnelle de la peine, auxquelles il nous faudra répondre.

Même si la comparaison entre mérite distributif et rétributif n'est pas une spécificité de ce chapitre, je vais indiquer brièvement pourquoi je pense qu'une analyse de la valeur communicationnelle du mérite en termes rétributifs est suffisante pour fournir une justification générale de la théorie communicationnelle du mérite. Tout d'abord, il existe un parallèle très clair entre les aspects communicatifs de la peine et de la récompense. De la même manière que la récompense peut servir à exprimer la gratitude et la reconnaissance de la société, la peine peut servir à exprimer la réprobation sociale. Dans ce cadre, il n'y a pas de raison de penser que la justification d'une théorie communicationnelle de la peine ne puisse servir également à justifier une théorie communicationnelle de la récompense.

Deuxièmement, il faut remarquer qu'une telle perspective n'est pas nouvelle. Dans sa *Théorie des sentiments moraux*, Adam Smith a effectué une défense communicationnelle de l'idée de mérite aussi bien dans sa perspective positive que négative¹⁹¹. Plus récemment, Axel Honneth a fourni une analyse des relations sociales sur le marché du travail sous l'angle de la reconnaissance¹⁹². Axel Honneth a défendu l'idée que la récompense du travail fournit une image de l'intégration de l'individu au sein de la communauté productive, qui constitue une part essentielle de la valeur de l'individu. Nous avons clairement ici affaire à une théorie communicationnelle du mérite distributif¹⁹³.

En plus de cette longue tradition philosophique, on peut également noter que des études sociologiques ont aussi défendu l'idée que les individus interprètent leur salaire sous un mode communicatif. Une telle perspective est très claire notamment dans le livre *Injustice*, dirigé par François Dubet, où un groupe de sociologues ont interrogé des centaines de salariés sur leur rapport au travail. Voici par exemple une des conclusions tirées par les auteurs : « C'est pour cette raison que si le mérite est associé à un sentiment d'accomplissement, l'échec est perçu comme une perte de dignité, de valeur et d'estime de soi, allant bien au-delà de ses conséquences

¹⁹¹ A. SMITH, *Théorie des sentiments moraux*, op. cit.

¹⁹² Axel HONNETH, *La lutte pour la reconnaissance*, trad. fr. Pierre RUSCH, Paris, Folio, 2013 ; Axel HONNETH, *Le droit de la liberté: Esquisse d'une éthicité démocratique*, trad. fr. Frédéric JOLY et trad. fr. Pierre RUSCH, Paris, Gallimard, 2015.

¹⁹³ Plus récemment, dans la tradition analytique, Dobos, dans son article « Exploitation, Working Poverty, and the Expressive Power of Wages », a effectué une analyse communicationnelle du tort effectué aux travailleurs sous-payés. Ned DOBOS, « Exploitation, Working Poverty, and the Expressive Power of Wages », *Journal of Applied Philosophy*, à paraître.

pratiques. On a honte d'avoir échoué dans les épreuves du mérite, et plus ces épreuves sont justes, plus on a honte¹⁹⁴. »

Bien sûr, une telle perspective montre une difficulté possible liée aux théories communicationnelles de la récompense. Si les récompenses sont interprétées comme indiquant la communication de l'appréciation du travail d'un individu, il semble clair qu'un salaire faible pourra être interprété comme un signe d'indignité. On peut s'inquiéter du climat social délétère qui pourrait être créé par une société qui chercherait à surinterpréter l'aspect communicatif du salariat¹⁹⁵. Ce problème se pose beaucoup moins dans le cas des théories communicationnelles de la peine, étant donné que la plupart des citoyens ne sont pas soumis à cette punition¹⁹⁶.

Malgré cette dernière difficulté, le parallèle entre théories communicationnelles de la peine et de la récompense est clair. Je me concentrerai donc essentiellement sur les débats beaucoup plus développés liés à la théorie communicationnelle de la peine. Ce chapitre procédera de la façon suivante. Dans un premier temps, je résumerai les grandes théories de justification de la peine, et j'indiquerai où la théorie communicationnelle de la peine se situe par rapport à elles. Dans un deuxième temps, je présenterai les différentes variantes des théories communicationnelles de la peine. Dans un troisième temps, j'indiquerai les principales critiques de ces théories. Une réponse à ces critiques nous occupera dans les chapitres 8 et 9.

Les grandes approches de la peine : rétribution, dissuasion, communication

La tradition philosophique a souligné la multiplicité des possibles justifications de la peine. Il est bien sûr impossible d'effectuer un panorama complet des différentes approches. Je présenterai ici simplement un bref résumé des grandes options possibles, afin que le lecteur puisse mieux saisir la spécificité des théories communicationnelles de la peine. Je me concentrerai sur trois grandes approches : les théories rétributivistes, dissuasives, et communicationnelles.

¹⁹⁴ F. DUBET, *Injustices. L'expérience des inégalités au travail*, op. cit., p. 96-97.

¹⁹⁵ Une telle perspective dystopique exploitant la valeur communicationnelle de la récompense se trouve dans l'ouvrage classique de Michael YOUNG, *The Rise of the Meritocracy 1870-2033*, Penguin Books, 1962.

¹⁹⁶ Étant donné la rareté des punitions, il existe un stigmatisme attaché à la peine qui peut conduire à l'opprobre social des personnes condamnées pour des fautes légères. On peut penser au cas littéraire de Jean Valjean, ou, de façon plus contemporaine, aux nombreuses peines de prison liées à la possession de cannabis ou liées aux infractions au code de la route.

Traditionnellement, on a divisé les approches en fonction du fait qu'elles étaient ou bien tournées vers le passé (*backward-looking*), vers l'acte que le criminel a fait, ou bien tournées vers l'avenir (*forward-looking*), vers la prévention de futurs crimes.

A un extrême, on trouve des théories purement tournées vers le passé, les théories rétributivistes brutes. Une théorie rétributiviste brute considère que la justification principale, voire unique, de la peine consiste dans le fait que faire souffrir quelqu'un en proportion du mal qu'il a causé est un bien en soi. Cette théorie se focalise donc uniquement vers le passé, c'est-à-dire vers l'acte commis par la personne ayant commis l'infraction. Il s'agit ici uniquement d'établir un type d'équilibre entre les torts commis par la personne condamnée et le mal que cette personne va elle-même subir. La théorie rétributiviste de la peine la plus connue est bien sûr celle développée par Kant. Une actualisation récente des théories rétributivistes de la peine peut être trouvée chez les philosophes Michael Moore ou Mitchell Berman¹⁹⁷.

A l'autre opposé, au sein des théories exclusivement tournées vers l'avenir, on trouve les théories dissuasives de la peine. Celles-ci considèrent uniquement les conséquences potentiellement bénéfiques de la peine, vis-à-vis de la prévention de crimes futurs. Faire souffrir la personne punie est considérée par de telles théories comme étant en soi injustifiée, sauf si cette souffrance sert à prévenir de futurs crimes. Les représentants les plus connus des théories dissuasives de la peine sont les théories utilitaristes, notamment celle proposée par Bentham. Plus récemment, des théories sophistiquées ont été proposées par Hart dans les années 60, et Braithwaite et Pettit plus récemment¹⁹⁸. On divise généralement les théories dissuasives de la peine en fonction de la manière spécifique par où procède la dissuasion : la dissuasion générale, la dissuasion spécifique, et l'incapacitation. La dissuasion générale vise à prévenir de futurs crimes en montrant au public les punitions qui attendent les criminels. La dissuasion spécifique vise à prévenir la récidive, en montrant au criminel ce qu'il encourt s'il commet une peine. L'incapacitation vise à écarter le criminel de la société, notamment via l'emprisonnement, afin qu'il soit incapable de commettre de nouveaux crimes pendant cette période. Ces trois fonctions de la peine sont, fondamentalement, réductibles à un seul but, la prévention de futurs crimes. En ce sens, il s'agit de trois variations sur un même thème dissuasif, exclusivement fondé sur les conséquences futures de la peine.

¹⁹⁷ Emmanuel KANT, *Métaphysique des moeurs : Tome 2, Doctrine du droit ; Doctrine de la vertu*, Flammarion, 2018 ; Michael S. MOORE, « Justifying Retributivism », *Israel Law Review*, 1993, vol. 27, n° 1-2, p. 15-49 ; Mitchell BERMAN, « Modest Retributivism », *Faculty Scholarship*, 2016.

¹⁹⁸ H. L. A. HART, *Punishment and Responsibility*, *op. cit.* ; John BRAITHWAITE et Philip PETTIT, *Not Just Deserts: A Republican Theory of Criminal Justice*, Oxford, Clarendon Press, 1993.

Comme nous l'avons vu dans le chapitre précédent, nous considérons que ces deux justifications constituent des justifications adéquates de la peine. Punir les criminels selon leur mérite contribue certainement à prévenir de futurs crimes, et en ce sens constitue une justification supplémentaire de cette pratique. De plus, punir les criminels de façon rétributive correspond à un processus extrêmement intuitif¹⁹⁹. En ce sens, en tant qu'intuition extrêmement ancrée, l'idée qu'il faut punir les individus selon leur mérite possède une forte plausibilité en tant qu'idée morale fondamentale.

Nous rejetons cependant une interprétation purement dissuasive de la peine, en tant qu'une telle interprétation nous semble inutilement violente et en rupture avec les droits fondamentaux des citoyens. De manière générale, il est plausible de penser que punir les criminels de la façon la plus dure possible sera la meilleure manière de faire peur aux citoyens et de punir des crimes futurs. Punir les citoyens au-delà de ce qu'ils méritent semblerait donc un bon moyen d'assurer la sécurité sociale. En ce sens, une théorie strictement dissuasive de la peine semble inacceptablement dure, et en rupture avec les intuitions courantes de proportionnalité entre le crime et la peine²⁰⁰.

Entre ces deux théories générales de la peine, basées sur la rétribution et la dissuasion, les théories communicationnelles de la peine ont eu une très grande influence au cours des quarante dernières années, aussi bien au sein des philosophes politiques qu'au sein des milieux juridiques. Selon les théories communicationnelles, le but de la peine est de communiquer la désapprobation sociale.

Les théories communicationnelles de la peine se présentent davantage sous la forme d'une famille de théories plutôt que comme une théorie unique. Une première division au sein des théories communicationnelles peut se comprendre en fonction du récipiendaire du message punitif. On voit là une division entre les théories dites expressives de la peine, d'un côté, et les théories qu'on peut appeler communicationnelles à proprement parler.

Selon les théories expressives de la peine, le but de la punition est de communiquer à la société que justice a été faite. La peine s'adresse donc aux autres citoyens, et non pas au criminel

¹⁹⁹ Un examen plus poussé des intuitions rétributives des individus aura lieu dans le chapitre suivant.

²⁰⁰ Bien sûr, des théories dissuasives sophistiquées ont été défendues, à la façon de celle proposée par Braithwaite et Pettit, qui considèrent que la peine ne doit pas servir simplement à minimiser les crimes futurs, mais à maximiser en général l'autonomie des citoyens. Maximiser l'autonomie des citoyens semblerait exclure l'idée de proposer des peines disproportionnées, parce que la peur d'être punie de façon barbare constituerait une limite claire à l'autonomie des citoyens. Même si de telles théories possèdent davantage de plausibilité que les théories dissuasives classiques, elles possèdent à leur tour de nombreuses difficultés. On peut notamment leur opposer l'idée qu'il n'est pas évident que la recherche de l'autonomie maximale ne conduirait pas à punir les criminels de façon maximale, en tant que ces derniers avaient déjà la possibilité d'éviter une telle peine, et ont donc exercé leur autonomie en choisissant une voie qui risquait de les faire punir. J. BRAITHWAITE et P. PETTIT, *Not Just Deserts*, *op. cit.*

lui-même. Une telle option théorique a une longue histoire, notamment depuis la théorie de la peine défendue par Emile Durkheim dans *De la division du travail social*, où Durkheim défendait l'idée que la peine visait à réaffirmer à la société la sacralité des valeurs sociales qui avaient été bafouées par le crime²⁰¹. Cette théorie a trouvé un renouveau récent au travers de l'article célèbre de Feinberg « The Expressive Function of Punishment ». Dans cet article, Feinberg défend que la fonction expressive de la peine sert trois objectifs différents. Tout d'abord, la peine constitue un désaveu public du criminel, symbolisant le fait que la société se désolidarise et met à l'écart le criminel. Ensuite, la peine constitue la critique publique du crime, et réaffirme l'opposition de la société vis-à-vis de cet acte. Troisièmement, la peine réactualise le caractère sacré de la loi²⁰². Toutes ces justifications présupposent que le punisseur s'adresse avant tout à la société dans son ensemble, plutôt qu'au criminel en particulier.

En opposition à cette idée d'une communication s'adressant à la société plutôt qu'au criminel, les théories communicatives de la peine conçoivent la peine comme l'instauration d'un dialogue entre le punisseur et la personne condamnée. Le but de la peine est alors de communiquer au criminel que la société et le punisseur désapprouvent ses actions, et condamnent ce qu'il a fait. Les théories communicationnelles de la peine se divisent encore en fonction de savoir si elles considèrent que le but de la peine est d'éduquer le condamné, ou si la communication est une fin en soi, sans qu'un objectif d'éducation morale soit également recherché. Parmi les théories qui considèrent que la peine n'a pas pour but d'éduquer en particulier le criminel, on peut prendre l'exemple de la théorie que Nozick a défendue dans le chapitre « Retributive punishment » de *Philosophical Explanations*²⁰³. Nozick y défend l'idée que le but de la peine est de mettre en lien le criminel et les valeurs que ce criminel a bafouées, en lui montrant que les valeurs ont bien un impact dans sa vie. Parmi les théories proprement éducatives de la peine, on peut prendre en exemple les théories récentes de Jean Hampton ou de Duff, qui considèrent que le but de la peine est d'offrir au condamné la possibilité de se repentir et de s'améliorer moralement²⁰⁴.

²⁰¹ Emile DURKHEIM, *De la division du travail social*, Paris, Presses Universitaires de France - PUF, 2013.

²⁰² J. FEINBERG, « The Expressive Function of Punishment », *op. cit.* Feinberg ajoute une quatrième fonction à l'expression de la peine : la proclamation de l'innocence des autres personnes qui auraient pu être suspectées. Il me semble moins évident de voir la nature du lien entre le fait d'innocenter une partie des suspects et la fonction expressive de la peine. On peut supposer que le fait d'innocenter aurait pu se faire de nombreuses manières différentes.

²⁰³ Robert NOZICK, « Retributive punishment », in *Philosophical Explanations*, Cambridge, Mass., Harvard University Press, 1981, p. 363-398.

²⁰⁴ Jean HAMPTON, « The Moral Education Theory of Punishment », *Philosophy & Public Affairs*, 1984, vol. 13, n° 3, p. 208-238 ; R. A. DUFF, *Punishment, Communication, and Community*, OUP, 2003. Duff refuse l'idée que sa théorie représente une théorie éducative de la peine. En effet, il considère que la peine conduit uniquement à la possibilité d'une amélioration morale des condamnés, et ne conduit pas avec certitude à une éducation morale.

Les théories communicatives de la peine forment donc un ensemble extrêmement vaste. Je voudrais ici défendre spécifiquement les théories éducatives de la peine, et non pas les théories dites expressives, ou les théories communicationnelles non-expressives. Il me semble en effet que ce sont les théories éducatives qui proposent la justification la plus convaincante du rôle expressif et symbolique de la punition.

Contre les théories expressives, où la peine s'adresse à la société et non pas à l'individu condamné, il est possible de mettre en avant le fait que la peine ne semblerait pas être foncièrement différente si elle était faite dans un cadre secret. Bien sûr, la perte de transparence serait dommageable, mais il ne semble pas qu'il serait absurde d'imaginer une société qui punirait dans un cadre privé afin de préserver l'anonymat de la victime et du condamné. De manière générale, si un procès se déroule dans l'indifférence générale, il ne semble pas que la punition s'en trouverait foncièrement gâchée. A l'inverse, un procès où le condamné n'apprendrait pas la raison pour laquelle il a été condamné semble à première vue absurde²⁰⁵.

En ce qui concerne les théories communicationnelles à proprement parler, justifier la punition par rapport à la possibilité d'éduquer un condamné semble plus plausible qu'une justification basée sur la volonté de communiquer par pur désir de communiquer. Si l'on savait *a priori* que la personne en question était complètement incapable de la moindre amélioration morale, il pourrait sans doute toujours y avoir un sens à chercher à la punir, afin de la dissuader de commettre des crimes à l'avenir, mais il serait moins intéressant de chercher à entrer dans un dialogue avec elle. Le but fondamental de tout dialogue semble quand même bien être de transmettre une information ou d'éduquer. A partir de maintenant, je référerai donc aux théories communicationnelles ou éducatives de la peine de façon indifférenciée.

Malgré le succès récent des théories communicationnelles de la peine, ces théories n'ont bien sûr pas été sans critiques. Avant de présenter ces dernières, je voudrais indiquer une nouvelle fois que cette justification générale de la peine ne doit évidemment pas se comprendre comme une justification de la situation pénale actuelle. Même si la pratique de la punition est justifiée en général, cela ne veut en aucun cas dire que toutes les pratiques pénales sont justifiées, ou que la surpopulation pénitentiaire ou l'incarcération de personnes pour des fautes relativement légères (comme des infractions au code de la route) puissent avoir un sens d'un

Jean Hampton n'avait pas dit autre chose quand elle avait développé sa théorie éducative de la peine. Je considère donc que la théorie de Duff est au fond très similaire à celle que Hampton avait développée, et j'utilise le terme général de théorie éducative pour les désigner toutes les deux.

²⁰⁵ Le chapitre suivant donnera plus de données empiriques pour nourrir cette intuition.

point de vue de la justice rétributive²⁰⁶. Les deux représentants majeurs des théories communicationnelles de la peine, Jean Hampton comme Anthony Duff, sont plutôt des réformateurs et considèrent que prendre en compte la peine comme communication implique de repenser de fond en compte le système pénal. Par exemple, Duff valorise des travaux d'intérêt général comme fournissant un modèle beaucoup plus sain de communication vis-à-vis des personnes condamnées. Il n'est pas évident que les travaux d'intérêt général puissent servir de modèle général pour une réforme pénale sans être accompagnés de peines plus dures. Cependant, le champ de possibilité de peines communicatives est ouvert, et ne saurait être limité à l'usage actuel massif de la prison.

La théorie communicationnelle de la peine : un nouveau type de dissuasion ?

Avant de passer aux critiques des théories communicationnelles de la peine, je souhaiterais mettre davantage en avant le lien entre théories communicationnelles, théories rétributives, et théories dissuasives. D'un côté, on pourrait avoir l'impression que les théories communicationnelles représentent un type de théories dissuasives, en tant que l'éducation du condamné peut apparaître comme une des conséquences possibles positives de la peine, conduisant à une réduction de crimes futurs. D'un autre côté, les théories communicationnelles ont souvent été interprétées comme des théories rétributives²⁰⁷. Les deux sentiments sont partiellement justifiés. En un sens, la théorie communicationnelle n'est pas vraiment un type de théorie rétributive, en tant qu'elle justifie en partie la peine par ses conséquences futures sur la possible éducation du condamné. Elle se rapproche en ce sens des théories dissuasives. Cependant, il ne s'agit pas non plus d'une théorie dissuasive, en tant que la théorie communicationnelle ne justifie pas la peine par la prévention de crimes futurs, mais directement par la possible éducation du condamné, sans prise en compte de conséquences ultérieures.

Il s'agit donc d'une justification intermédiaire entre rétribution et dissuasion. Elle ressemble aux théories accordant une valeur intrinsèque à la peine en ce que le fait de communiquer n'est

²⁰⁶ Pour un examen de certaines aberrations des politiques pénales actuelles, on pourra consulter le livre ethnographique de Didier FASSIN, *L'Ombre du monde. Une anthropologie de la condition carcérale*, Points, 2017.

²⁰⁷ Par exemple, Nozick présente sa théorie communicationnelle comme un type de théorie rétributive dans « Retributive punishment », *op. cit.*

pas une conséquence ultérieure, mais une conséquence intrinsèquement liée au fait de récompenser ou de punir quelqu'un. On peut mettre en avant que la punition n'est pas simplement un instrument de la communication, mais fait partie de la communication elle-même. Pour les théories communicationnelles, la peine est le meilleur instrument pour communiquer au condamné que son action était mauvaise. En ce sens, les théories communicationnelles justifient la punition elle-même, alors que les théories dissuasives justifient uniquement ses conséquences lointaines. De manière plus précise, pour les théories communicationnelles, punir n'est pas « en soi » bon ou moralement bon, mais est en pratique toujours bon parce qu'il existe un lien empirique constant entre punition et communication.

Les théories communicationnelles de la peine trouvent donc une bonne partie de leur originalité dans leur insistance sur le fait que la peine constitue un moyen privilégié de communication, au-delà de la simple communication orale. C'est sur ce point que se sont concentrées la grande majorité des critiques.

Les critiques traditionnelles de la théorie communicationnelle de la peine

Différentes critiques ont été fournies contre la théorie communicationnelle de la peine. Une critique précoce, formulée avant que les théories communicationnelles de la peine reviennent à la mode à partir des années 1960-1970, a été formulée par Hart en 1963 dans son *Law, Liberty, and Morality*, où Hart s'opposait aux théories moralisatrices de la loi, dans un contexte de débat sur la dépénalisation de l'homosexualité en Grande-Bretagne. Ces critiques ont été renouvelées au cours des ans, notamment par des auteurs comme T. M. Scanlon, Russ Shafer-Landau, ou Nathan Hanna²⁰⁸. Je vais rappeler les deux principales critiques, qui se retrouvent chez presque tous ces opposants.

Tout d'abord, les critiques ont avancé qu'il est barbare de chercher à faire souffrir quelqu'un afin de lui communiquer quelque chose. La souffrance d'un individu ne peut jamais être justifiée par le fait que l'on essaye de lui transmettre une information. Ensuite, il a été reproché

²⁰⁸ H. L. A. HART, *Law, Liberty and Morality*, Oxford University Press, Oxford, 1968 ; T. M. SCANLON, « The Significance of Choice », in Sterling M. MCMURRIN (dir.), *The Tanner Lectures on Human Values (Vol. 8, pp. 149-216)*, University of Utah Press, 1988, p. ; Russ SHAFER-LANDAU, « Can Punishment Morally Educate? », *Law and Philosophy*, 1991, vol. 10, n° 2, p. 189-219 ; Nathan HANNA, « Say What? A Critique of Expressive Retributivism », *Law and Philosophy*, 2008, vol. 27, n° 2, p. 123-150.

à la théorie communicationnelle de la peine d'être inutile : si le seul but est de communiquer quelque chose, il serait plus efficace d'user simplement de la parole, sans en venir à utiliser un moyen physique. Ces deux critiques se retrouvent de façon particulièrement éloquente dans le passage suivant de Hart : « L'idée que nous pouvons punir des personnes ayant violé un code moral, non pas pour prévenir des dommages ou de la souffrance ou même la répétition de l'offense, mais simplement en tant que moyen d'exprimer de façon emphatique notre condamnation morale, s'approche dangereusement du sacrifice humain en tant qu'expression d'adoration humaine. Mais même si nous rejetons cette objection, une seconde s'offre à nous. Qu'est-ce que nous voulons dire quand nous disons que la punition des délinquants est une manière appropriée d'exprimer une condamnation morale emphatique ? Le moyen habituel d'exprimer la condamnation morale passe par des mots, et ce n'est pas évident, si une dénonciation est vraiment requise, pourquoi une déclaration publique solennelle ne serait pas la manière la plus « appropriée » ou « emphatique » de l'exprimer. Pourquoi est-ce qu'une dénonciation devrait prendre la forme d'une punition²⁰⁹ ? »

Même s'il s'agit, dans leur fond, des deux critiques les plus puissantes des théories communicationnelles de la peine, il existe d'autres contre-arguments intéressants. On trouve par exemple dans l'article de Russ Shafer-Landau « Can punishment morally educate ? » un renforcement des critiques contre l'usage de la souffrance pour éduquer quelqu'un, dans le cadre d'une attaque contre les contradictions internes aux théories éducatives²¹⁰. Russ Shafer-Landau montre que de nombreuses théories éducatrices classiques, comme celles défendues par Morris ou Jean Hampton, sont des théories qui cherchent à promouvoir le bien du condamné. L'éducation morale correspondrait à un besoin du délinquant, et la peine serait alors la réponse à ce besoin de l'âme. Shafer-Landau met en avant de nombreux problèmes avec l'idée qu'il faille faire souffrir quelqu'un pour son bien. Par exemple, Shafer-Landau met en avant que la punition représente une violation de l'autonomie de quelqu'un ; il est difficile de penser pourquoi une telle violation de l'autonomie peut être justifiée dans ce cas particulier, alors qu'elle ne l'est généralement pas. Ensuite, plus grave, Shafer-Landau souligne que, si l'on veut dire que l'éducation morale représente un bien pour le condamné, alors il faut développer une théorie des intérêts qui mette en avant pourquoi le fait d'être égoïste représenterait un mal pour

²⁰⁹ H. L. A. HART, *Law, Liberty and Morality*, *op. cit.*, p. 65-66. Ces deux critiques se retrouvent chez presque tous les auteurs, et se retrouvent même chez des défenseurs de la théorie communicative de la peine, comme Nozick. R. NOZICK, « Retributive punishment », *op. cit.*, p. 374.

²¹⁰ R. SHAFER-LANDAU, « Can Punishment Morally Educate? », *op. cit.*

la personne concernée. Faute d'avoir développé cette théorie, il est douteux que l'on puisse dire que faire souffrir une personne puisse être fait pour son propre bien.

Enfin, on peut mettre en avant une quatrième grande critique contre les théories communicatives de la peine. Dans une série de travaux récents, le sociologue Didier Fassin a reproché aux théories de la peine de ne pas correspondre réellement aux intuitions des gens, et d'être un simple vernis idéologique visant à justifier des intuitions en réalité strictement basées sur un désir de faire souffrir les délinquants²¹¹. Cette critique est très générale ; elle peut valoir en effet contre toute théorie de la peine. Cependant, il me semble qu'elle possède une force particulière dans le cadre des théories communicatives. La théorie communicationnelle de la peine perdrait une grande partie de sa valeur si les gens n'interprètent pas la punition ou la récompense en fait comme des formes de communication, et s'ils ne pensent pas qu'il s'agit de quelque chose d'important. Les théories communicationnelles de la peine sont fondées sur l'idée que la punition constitue un moment éducatif. Mais la communication, tout comme l'éducation, peuvent difficilement se passer de l'adhésion des acteurs à l'acte punitif. Une communication inintentionnelle, ou une éducation sans éducateur, semblent difficile à concevoir. Pour que la théorie communicationnelle de la peine fasse sens, il faut que les acteurs eux-mêmes perçoivent l'éducation comme un objectif important.

Conclusion

Dans ce chapitre, nous avons commencé à présenter les différentes théories de la peine, et plus particulièrement les théories dites communicationnelles. J'ai indiqué que je favorisais personnellement les théories qui considèrent que l'éducation est un des buts possibles de la peine. Nous avons vu dans un second temps quatre grandes critiques de la peine. Ces critiques reprochaient aux théories communicationnelles d'être barbares en justifiant la souffrance des condamnés, d'être hypocrites en prétendant agir pour le bien du condamné, d'être obscures en cherchant à fournir une communication en dehors des outils linguistiques habituels, et d'être enfin déconnectées des intuitions et des pratiques

Ces critiques possèdent une certaine force, et j'y répondrai de façon détaillée dans les deux prochains chapitres. Dans un premier temps, dans le chapitre 8, je commencerai par examiner

²¹¹ Didier FASSIN, *Punir. Une passion contemporaine*, Paris, Le Seuil, 2017.

les données empiriques relatives à la perception de la valeur communicative de la peine. J'y démontrerai que la communication joue effectivement un rôle important dans la perception des pratiques punitives. Cela nous permettra de répondre aux critiques considérant que les théories communicationnelles sont complètement déconnectées des pratiques effectives. De plus, suivant les idées de justification développées dans les chapitres précédents, cet accord entre la théorie communicationnelle de la peine et les intuitions courantes fournira une justification *prima facie* à cette théorie.

Enfin, dans un deuxième temps, dans le chapitre 9, je répondrai aux autres critiques théoriques. Je défendrai ainsi l'idée qu'infliger intentionnellement de la souffrance n'est pas nécessairement barbare, et, surtout, je fournirai une justification générale de l'idée que la punition peut être un outil utile dans le cadre d'une stratégie communicative.

Chapitre 8

La valeur communicative de la peine : données empiriques

Dans le chapitre précédent, nous avons vu un ensemble d'arguments tendant à montrer que la valeur communicative de la récompense et de la peine constituait une justification possible des pratiques de mérite. De tels arguments, cependant, ne possèdent qu'une valeur très limitée tant qu'il n'est pas démontré que l'idée de punition comme communication possède un fondement empirique.

De fait, si les pratiques punitives réelles se font sans que la punition ne soit interprétée sous un mode communicatif, la théorie communicative de la peine perdrait une grande partie de sa plausibilité. En effet, dans une telle situation, les partisans d'une théorie communicative de la peine devraient défendre une théorie révisionniste de nos pratiques punitives, en défendant l'idée que nos pratiques devraient développer un aspect communicatif qui est actuellement absent. Il s'agirait alors de changer notre psychologie afin que nous soyons davantage sensibles aux aspects communicatifs de la punition.

Ce chapitre procédera principalement en deux temps. Dans un premier temps, j'effectuerai une brève présentation des données empiriques quant au lien psychologique entre punition et communication. Ensuite, dans un deuxième temps, je présenterai les résultats de trois expériences que j'ai effectuées avec Geoff Goodwin, de l'université de Pennsylvanie (*UPenn*), et avec Paulo Sousa, de la Queen's University Belfast (*QUB*). L'ensemble de ces expériences permettra de démontrer l'importance de la communication dans la perception de la justice de la peine, et dans la conceptualisation de la punition.

Il existe une multiplicité de questions possibles qui peuvent conduire à un examen du lien entre communication et punition. On peut mettre en avant cinq questions principales.

- 1) *La question de la légitimité.* Tout d'abord, on peut se demander dans quelle mesure une punition apparaît comme davantage juste et légitime si elle est effectuée d'une manière qui permet une pleine communication au délinquant, et si le délinquant comprend bien les raisons pour lesquelles il a été puni.

- 2) *La question conceptuelle.* Cette deuxième question consiste à étudier dans quelle mesure la punition prototypique inclut l'idée d'une communication.
- 3) *La question motivationnelle.* On passe à un niveau comportemental : il s'agit de savoir si les gens sont davantage motivés à punir s'ils ont la possibilité d'instaurer une communication avec le délinquant.
- 4) *La question de l'interprétation.* Il s'agit ici de savoir si les gens interprètent les actes effectifs de punition comme possédant une valeur symbolique, expressive ou communicative.
- 5) *La question du succès éducatif.* Est-ce que les personnes punies ont effectivement plus de chances de passer par un processus de remords et d'éducation morale que les personnes qui n'ont pas été punies ?

Pour établir le plein succès d'une théorie communicationnelle, il faudrait qu'il soit possible de montrer un impact de la communication sur tous ces problèmes différents. Bien sûr, il s'agit d'un programme de recherche gigantesque, et je ne prétends pas apporter une réponse définitive à toutes ces questions. Dans le cas du succès éducatif, la recherche manque, et je ne prétendrai même pas apporter un début de réponse. Pour les quatre premiers points, cependant, je chercherai à montrer qu'il est possible d'indiquer que la communication a effectivement un impact.

La psychologie de la peine : un clair soutien au rétributivisme

Avant de présenter les recherches récentes sur le lien entre communication et punition, je souhaiterais indiquer brièvement un résultat extrêmement net en psychologie de la peine, à savoir, le fait que les intuitions populaires sont beaucoup plus proches des philosophies rétributives que des philosophies dissuasives. Au sein d'un vaste ensemble de recherches, je présenterai deux types d'expériences qui ont contribué à montrer l'importance du rétributivisme populaire.

En 1982, Robert M. McFatter a interrogé 72 étudiants sur leur perception de la punition liée à un ensemble de crimes incluant une agression, un vol de voiture, un viol, et un meurtre²¹². McFatter a assigné les participants à 6 conditions, assignant chaque groupe de 12 étudiants à

²¹² Robert M. MCFATTER, « Purposes of punishment: Effects of utilities of criminal sanctions on perceived appropriateness », *Journal of Applied Psychology*, 1982, vol. 67, n° 3, p. 255-267.

étudier une perspective pénale particulière. 5 groupes devaient adopter ou bien la perspective d'un défenseur de la dissuasion générale, de la dissuasion spécifique, de l'incapacitation, de la réhabilitation, ou de la rétribution, et devaient indiquer sur une échelle de 1 à 10 dans quelle mesure différents types de peine (allant d'un mois de prison à la peine de mort) seraient appropriés en fonction de la perspective philosophique qu'on leur avait demandé d'adopter. Enfin, le 6^e groupe devait simplement indiquer quelle peine lui semblait être la plus appropriée pour chaque type de crime. McFatter a découvert que les peines recommandées par les étudiants assignés à la position rétributive étaient très proches des peines préconisées par les étudiants qui devaient simplement indiquer la peine qui leur semblait la plus appropriée en général²¹³. A l'inverse, adopter les différentes positions dissuasives conduisait les étudiants à recommander des peines beaucoup plus lourdes ; de fait, les étudiants indiquaient que la peine qui semblerait la plus efficace d'un point de vue purement dissuasif était la peine de mort, ou la prison à vie, pour n'importe quel type de crime. On voit ici un soutien empirique à l'argument philosophique contre les théories dissuasives, fondé sur l'idée que les théories dissuasives conduisent à une philosophie pénale beaucoup trop sévère.

Plus récemment, deux articles ont montré que les participants américains possèdent des intuitions rétributivistes, même en l'absence de toute possible communication, et donc *a fortiori* de tout effet dissuasif. Goodwin et Benforado ont ainsi montré que les gens possèdent des intuitions rétributivistes même vis-à-vis des animaux²¹⁴. Dans une toute autre expérience, Nadelhoffer et ses collègues ont montré que les américains étaient susceptibles de punir des individus ayant pris de l'argent à une tierce personne, et ce, même dans une situation où cet individu ne savait pas qu'il allait être puni²¹⁵. Ces expériences ont montré que la psychologie de la punition possède un fondement de rétributivisme brut : les individus sont prêts à infliger un dommage à quelqu'un, et considèrent qu'il s'agit d'un bien en soi, même dans des circonstances où il n'y a pas de possibilité de communication.

Je vais rapidement présenter les expériences de Goodwin et Benforado où ces derniers ont montré l'existence d'intuitions rétributives vis-à-vis d'animaux. La punition des animaux possède une longue histoire dans les sociétés humaines, et était notamment assez fréquente dans les sociétés du Moyen-Âge occidental. L'idée à la base de leurs expériences est que la punition

²¹³ Les peines recommandées par les étudiants assignés à la perspective réhabilitatrice étaient également proches de la position rétributive.

²¹⁴ Geoffrey P. GOODWIN et Adam BENFORADO, « Judging the Goring Ox: Retribution Directed Toward Animals », *Cognitive Science*, 2015, vol. 39, n° 3, p. 619-646.

²¹⁵ Thomas NADELHOFFER, Saeideh HESHMATI, Deanna KAPLAN et Shaun NICHOLS, « Folk Retributivism and the Communication Confound », *Economics & Philosophy*, 2013, vol. 29, n° 2, p. 235-261.

vis-à-vis d'un animal ayant blessé un être humain peut s'expliquer de deux manières : ou bien par des intuitions rétributives, ou bien parce que l'animal, étant dangereux, doit être mis hors d'état de nuire à l'avenir. Il existe donc une interprétation rétributiviste, et une interprétation dissuasive. Pour départager ces deux interprétations, Goodwin et Benforado ont présenté à 500 participants recrutés sur *Amazon Mechanical Turk* différents meurtres causés par un requin²¹⁶. La stratégie des auteurs reposait sur une variation dans l'identité des victimes : selon les vignettes, le requin avait tué un pédophile de 48 ans, ou une fillette de 10 ans. L'idée des auteurs était que l'animal devrait apparaître comme également dangereux dans les deux cas. Si la perception de la dangerosité de l'animal variait selon les vignettes, cette perception devrait aller dans le sens d'une plus grande dangerosité dans le cas d'un animal capable de tuer un homme adulte plutôt qu'une enfant de 10 ans. Pour autant, étant donné que les individus accordent en général plus d'importance à la valeur de la vie des enfants que des personnes de 50 ans, ils devraient ressentir plus d'indignation dans le cas du meurtre de la fillette ; la pédophilie devrait, de plus, renforcer ce sentiment d'une moindre valeur pour l'homme pédophile²¹⁷. Si les gens ont des intuitions rétributives, ils devraient donc vouloir punir un animal plus fortement pour le meurtre d'une fillette que pour le meurtre de l'homme de 48 ans. C'est exactement ce que les auteurs ont trouvé : les participants avaient beaucoup plus tendance à déclarer être favorables au fait que les autorités mettent l'animal à mort, et beaucoup plus tendance à dire que l'animal méritait d'être mis à mort, dans le cas de la fillette. De plus, les participants avaient beaucoup plus tendance à dire que l'animal devrait ressentir de la douleur lors de la mise à mort, et qu'il n'était pas besoin de se servir d'un moyen indolore pour le tuer.

Ces expériences montrent bien que les individus possèdent des intuitions qui sont globalement conformes au rétributivisme, et qu'ils peuvent soutenir des mesures punitives même dans des situations où il est impossible de communiquer quoi que ce soit. Cependant, le fait que les individus possèdent des intuitions en partie fondées sur un rétributivisme brut ne prouve pas que leurs intuitions vis-à-vis de la punition ne sont pas pluralistes. Les intuitions populaires pourraient se fonder sur une pluralité de justifications. Il s'agira de prouver ici que les intuitions populaires se fondent également sur une justification communicative, et qu'elles considèrent qu'il est important de punir en partie parce que cela conduit à communiquer une

²¹⁶ G. P. GOODWIN et A. BENFORADO, « Judging the Goring Ox », *op. cit.* Les données sont celles de l'étude 5, qui a été sélectionnée parce qu'il s'agit de la plus convaincante des différentes études utilisées, mais l'ensemble des expériences présentent des résultats allant dans la même direction.

²¹⁷ Le fait que les individus accordent en général plus de valeur à la vie d'un enfant qu'à la vie d'une personne adulte a bien été démontré dans un autre article, écrit par Geoff Goodwin et Justin Landy. Geoffrey P. GOODWIN et Justin F. LANDY, « Valuing different human lives », *Journal of Experimental Psychology. General*, 2014, vol. 143, n° 2, p. 778-803.

réprobation sociale au criminel. Il est temps maintenant d'examiner une série d'expériences récentes qui ont tenté de prouver l'existence d'intuitions conformes aux prédictions d'une théorie communicative de la peine²¹⁸.

La valeur expressive de la loi

Un test récent de la question de savoir dans quelle mesure les gens interprètent la punition comme communiquant quelque chose sur la valeur des délinquants a été effectué par Kenworthy Bilz dans son article « Testing the Expressive Theory of Punishment²¹⁹ ». Dans cet article, Bilz a présenté une série d'expériences afin de déterminer dans quelle mesure des individus faisaient des inférences sur le statut social d'une personne punie. Pour l'auteur, si la peine communique la réprobation sociale, alors la punition devrait affecter le statut de la personne condamnée.

Pour tester cette hypothèse, Bilz a présenté à un assez grand nombre de participants (avec des échantillons allant jusqu'à 300 participants) des scénarios où un ou plusieurs personnes étaient accusées de divers crimes ou délits. Ces crimes et délits incluaient notamment un cas de viol, et un cas de vol d'identité. La manipulation expérimentale consistait à faire varier le type de punition infligée à la personne ayant commis le délit ou le crime. Dans une condition, la peine était légère, dans l'autre beaucoup plus conséquente. Les variables dépendantes de

²¹⁸ Parmi tous les articles que j'aurais pu présenter sur la théorie communicative de la peine, j'ai choisi de ne pas faire de présentation des articles effectués par Mario Gollwitzer et son équipe. Ces expériences visaient à montrer que les gens sont beaucoup plus satisfaits quand ils ont eu la possibilité d'envoyer un message à la personne qu'ils ont punie. Dans toutes les expériences dont s'est servi Gollwitzer, le punisseur était la victime de l'acte ; autrement dit, il ne s'agissait pas d'une punition impartiale, comme dans la plupart des modèles philosophiques de la peine. Je ne parlerai pas non plus des expériences faites par Fieri Cushman et son équipe. Cushman est un psychologue important qui a défendu de façon très forte l'idée que la punition a un but éducatif. Cependant, il n'a effectué aucune expérience testant directement cette hypothèse, et a surtout mis l'accent sur l'existence de preuves indirectes. Par exemple, Cushman interprète l'existence de cas de fortune morale comme révélant le fait que nous nous saisissons de toutes les possibilités pour éduquer moralement nos concitoyens, même dans des cas ambigus. Il ne s'agit pas cependant d'une preuve très forte de l'existence d'une logique éducative ou communicationnelle dans les pratiques pénales. Mario GOLLWITZER et Markus DENZLER, « What Makes Revenge Satisfactory: Seeing the Offender Suffer or Delivering a Message? », *Journal of Experimental Social Psychology*, 2009, vol. 45, n° 4, p. 840-844 ; Mario GOLLWITZER et Milena MEDER, « What gives victims satisfaction when they seek revenge? », *European Journal of Social Psychology*, 2011, vol. 41, n° 3, p. 364-374 ; F. FUNK, V. MCGEER et M. GOLLWITZER, « Get the Message: Punishment Is Satisfying If the Transgressor Responds to Its Communicative Intent », *Personality and Social Psychology Bulletin*, 2014, vol. 40, n° 8, p. 986-997 ; Fieri CUSHMAN, Anna DREBER, Ying WANG et Jay COSTA, « Accidental Outcomes Guide Punishment in a "Trembling Hand" Game », *PLoS ONE*, 2009, vol. 4, n° 8.

²¹⁹ Kenworthy BILZ, « Testing the Expressive Theory of Punishment », *Journal of Empirical Legal Studies*, 2016, vol. 13, n° 2, p. 358-392.

l'expérience portaient sur la perception de la personne condamnée. Bilz a demandé aux participants comment les habitants de l'endroit où s'est passée l'affaire percevraient le condamné en termes d'admiration, de valorisation, et de respect.

Bilz a obtenu des résultats très nets. La personne qui avait été punie sévèrement était notée de façon dure, indiquant que les sujets considéraient qu'elle devait être beaucoup moins admirée, valorisée, respectée, que la personne qui avait été punie de façon moins sévère. Et ce, alors même que les faits reprochés étaient exactement les mêmes dans les deux cas. Ce genre d'expérience indique clairement que la peine est perçue comme un stigmate social, et qu'une personne punie est perçue comme possédant un statut moindre qu'une personne qui a été légèrement punie. Il existe des conséquences symboliques fortes de la peine. Les personnes interprètent donc la peine sur un mode expressif.

Cette valeur expressive de la punition a été confirmée dans une autre expérience récente par Jessica L. Bregant et ses collègues²²⁰. Alors que Bilz s'est concentrée sur ce que la punition communique à propos du criminel, Bregant et ses collègues ont étudié ce que la punition exprime vis-à-vis du crime. Dans une série d'études, Bregant et ses collègues ont présenté à plus d'un millier de participants recrutés sur *Amazon Mechanical Turk* divers scénarios imaginaires présentant une population extraterrestre. Cette population était décrite comme similaire à celle des êtres humains. Les chercheurs ont alors présenté aux participants deux actions aux noms inventés : « Blicking » et « gomping ». Les participants apprenaient alors que l'action de « blicking » était punie, tandis que « gomping » ne l'était pas. Les participants devaient alors indiquer laquelle des deux actions était pire d'un point de vue moral. De façon assez peu surprenante, 95% des participants ont indiqué que l'action punie était pire d'un point de vue moral que l'action qui n'était pas punie par la loi. En cas de situation ambiguë, les personnes déduisent donc qu'une action punie est probablement pire qu'une action qui ne l'est pas.

Ces deux expériences montrent bien que la loi exprime quelque chose, et que les individus font des inférences de la punition à la valeur de l'action. Pour autant, ce genre d'expériences ne permet pas de déterminer dans quelle mesure les individus punissent pour exprimer quelque chose, pour transmettre l'idée de blâme. Il est tout à fait possible que la dimension symbolique de la peine soit une simple conséquence, qui n'intervienne pas du tout dans les motivations individuelles.

²²⁰ Jessica BREGANT, Eugene M. CARUSO et Alex SHAW, *Expressing What? Evaluating the Expressive Value of Punishment*, Rochester, NY, Social Science Research Network, 2016.

Nous allons donc maintenant examiner une expérience récente qui a cherché à montrer que la punition a un but uniquement rétributif, et que la communication ne motive pas les individus à punir davantage. Une telle expérience représenterait un défi important pour les théories communicationnelles de la peine.

Contre l'importance de la communication dans la punition : l'expérience de Nadelhoffer

Je vais présenter une expérience importante faite par Nadelhoffer et ses collègues en 2014 et publiée sous le titre « Folk Retributivism and the Communicative Confound²²¹ ». Étant donné que notre première étude, présentée plus loin, constitue une réplique conceptuelle de cette expérience, je vais davantage rentrer dans les détails de cette étude.

Dans leur expérience, Nadelhoffer et ses collègues ont recruté 74 participants sur *Amazon Mechanical Turk*. Nadelhoffer a alors demandé aux participants d'imaginer qu'ils prenaient part à un jeu économique avec deux autres personnes²²². Une première personne, appelée G, devait recevoir 2 dollars. Une deuxième personne, appelée D, avait la possibilité de faire disparaître ces 2 dollars. En d'autres termes, D avait la possibilité d'empêcher la première personne de recevoir l'argent, sans pour autant recevoir elle-même davantage d'argent. Le sujet devait s'imaginer comme étant une troisième personne, qui devait indiquer si elle voudrait perdre de l'argent à son tour pour retirer de l'argent à la deuxième personne au cas où cette deuxième personne avait fait le choix de retirer de l'argent à la première personne. Autrement dit, même si les instructions n'employaient pas ces mots, le participant devait s'imaginer comme étant une sorte de juge, qui avait la possibilité de punir quelqu'un qui avait effectué une action anti-sociale. La personne jouant le rôle du « juge » avait trois choix : ou bien ne pas punir, auquel cas la deuxième personne (D) et le participant recevraient 4 dollars ; ou bien le participant pouvait payer 1 dollar (et donc n'en recevoir que 3) pour enlever 3 dollars à la

²²¹ T. NADELHOFFER, S. HESHMATI, D. KAPLAN et S. NICHOLS, « Folk Retributivism and the Communication Confound », *op. cit.*

²²² Un jeu économique est un mode expérimental courant en psychologie sociale ou en économie comportementale. Ces expériences consistent à proposer aux participants différents choix, qui correspondent à différentes récompenses monétaires. Un jeu économique célèbre est par exemple le jeu du dictateur, où deux participants doivent se partager une certaine somme.

personne punie ; ou bien le participant pouvait payer 2 dollars pour que la personne punie ne reçoive rien du tout.

Le participant pouvait être mis dans deux conditions. Dans une première condition, la condition *Connaissance*, il était dit aux participants que le deuxième participant savait qu'il pouvait recevoir un bonus pouvant aller jusqu'à 4 dollars, et qu'il savait que le montant de ce bonus dépendait du comportement d'un troisième acteur. Dans la deuxième condition, la condition *Ignorance*, il était dit aux participants que la deuxième personne ne savait pas du tout qu'elle pouvait recevoir un bonus allant jusqu'à 4 dollars, et qu'elle ne saurait pas dans quelle mesure de l'argent avait été retiré.

Le raisonnement des expérimentateurs était le suivant : dans le cas de la condition *Ignorance*, il est difficile de voir comment la punition pourrait résulter de motifs communicationnels. En effet, la personne punie n'aurait aucun moyen d'inférer dans quelle mesure on lui avait retiré de l'argent par rapport à ce qu'il aurait dû recevoir. A l'inverse, dans le cadre de la condition *Connaissance*, le motif communicatif pouvait s'exprimer, puisque la personne punie savait qu'elle devait recevoir de l'argent. Une proportion plus forte de personnes choisissant de punir dans le cadre de la condition *Connaissance* impliquerait que la communication a bien un rôle dans les motivations derrière la punition.

En l'occurrence, le résultat a été négatif : ils ont trouvé des résultats similaires dans les deux conditions, avec 69% des gens choisissant de retirer de l'argent dans la condition *Ignorance*, et 57% choisissant de retirer de l'argent dans la condition *Connaissance*. Contrairement aux prédictions de la théorie communicative de la peine, il y a eu moins de participants choisissant de punir dans la condition avec une possibilité communicative, même si cette différence n'est pas significative, et qu'on peut raisonnablement penser qu'elle est uniquement due à des variations aléatoires.

Les expérimentateurs concluent que leurs résultats sont compatibles avec un rétributivisme brut. Dans la punition, il n'y a pas d'objectif dissuasif ou d'objectif communicationnel : il s'agit simplement d'un désir de faire souffrir, ou au moins d'infliger quelque chose de négatif, à une personne qui a agi négativement. Il n'y aurait pas d'autre objectif derrière la punition.

Une telle conclusion est particulièrement surprenante. En effet, toute l'histoire de la philosophie pénale et des débats politiques sur la punition possède une forte imprégnation de motifs à visée éducative. On peut considérer qu'il s'agit d'une pure idéologie, mais il semblerait toutefois peu probable que la possibilité de communiquer n'ait aucun impact sur la volonté de punir.

Cet appel à l'éducation dans les discours punitifs n'est d'ailleurs pas simplement une spécificité occidentale ; on la retrouve dans d'autres cultures. Par exemple, dans son travail de terrain chez les Turkana de Tanzanie, Sarah Mathew rapporte que les recommandations punitives se font régulièrement en faisant appel à la valeur éducative de la peine. Elle cite par exemple un Turkana qui lui a déclaré : « Pour lui apprendre, il faut le battre, et pas seulement lui parler. [...] C'est quand il sentira la douleur qu'il regrettera que son peuple l'ait maudit²²³ [...] »

Dans ce contexte, nous avons décidé de répliquer l'expérience de Nadelhoffer et de ses collègues, avec un modèle expérimental qui met davantage en avant la valeur communicative de la punition. De plus, les expériences citées plus haut ne permettent de répondre que très partiellement aux interrogations que l'on peut se poser sur le lien entre communication et punition. Ces expériences ont montré que les gens interprètent bien la punition sur un mode expressif, et qu'ils ne sont pas davantage motivés à punir quand ils ont la possibilité de communiquer. Cependant, cela ne donne aucune information sur la perception de la légitimité de la punition à visée communicationnelle, et cela n'indique pas non plus dans quelle mesure les personnes considèrent que le concept de punition inclut l'idée de communication. Nous remédions à ces limitations dans nos trois expériences.

Nouvelles expériences sur la théorie communicationnelle de la peine

Les sections précédentes ont commencé à présenter plusieurs résultats empiriques sur les théories communicationnelles de la peine. Nous avons notamment vu, à partir des travaux de Kenworthy Bilz et de Jessica Bragant, que les individus interrogés interprètent la punition comme révélant le statut social de la personne condamnée, et comme exprimant la gravité de l'offense en question. Nous avons également vu que les expériences conduites par Thomas Nadelhoffer tendent à montrer que la volonté de communiquer semble avoir peu d'impact sur la motivation effective des individus à punir. Un tel résultat semble aller à l'encontre des prédictions des théories communicationnelles de la peine.

Dans les prochaines sections, je vais présenter les résultats de trois expériences conduites avec Geoff Goodwin et Paulo Sousa. La première expérience constitue une réplique

²²³ Sarah MATHEW et Robert BOYD, « The cost of cowardice: punitive sentiments towards free riders in Turkana raids », *Evolution and Human Behavior*, 2014, vol. 35, n° 1, p. 58-64.

extension de l'expérience de Nadelhoffer. Elle vise à tester de façon plus approfondie le résultat de Nadelhoffer selon lequel la communication n'a pas d'impact motivationnel. Comme je le montrerai, nos résultats sont plus ambigus que ceux de Nadelhoffer : la possibilité de communiquer semble constituer une incitation à punir, mais uniquement si l'importance de la communication a été rendue saillante.

Nos expériences 2 et 3 visent à tester une tout autre hypothèse. Passant outre la question de la motivation, elles visent à étudier dans quelle mesure une punition est perçue comme étant plus légitime si elle possède un aspect communicationnel, et dans quelle mesure la communication appartient au concept de punition. Par rapport à la question de légitimité, nous trouvons que les participants considèrent qu'une punition est beaucoup plus légitime quand elle possède un aspect communicatif. Par rapport à la question conceptuelle, nos résultats tendent à montrer que la communication appartient à l'idée d'une punition prototypique, même si la communication ne semble pas être une condition nécessaire de l'application du terme « punition ».

Expérience 1

Notre *Expérience 1* est une réplique relativement proche de l'expérience de Nadelhoffer et de ses collègues. Nous avons décidé de répliquer leur expérience pour plusieurs raisons. Tout d'abord, leur échantillon était assez faible. Ils avaient recruté 74 participants, ce qui est peu, surtout dans le cadre d'une expérience qui conclut à un résultat négatif (« absence de preuve n'est pas preuve d'absence », comme dit le dicton). Nous avons recruté 232 participants et fait usage d'un plan expérimental croisé (*within-subject design*), qui possède une plus grande puissance statistique que le plan expérimental emboîté (*between-subject design*) utilisé par Nadelhoffer et ses collègues²²⁴.

Ensuite, d'autres facteurs, liés spécifiquement au contenu des instructions distribuées par Nadelhoffer, nous ont conduit à vouloir répliquer leur expérience. Premièrement, nous avons voulu clarifier l'aspect communicatif de l'expérience, en permettant l'envoi d'un message dans

²²⁴ Un plan expérimental croisé conduit tous les participants à effectuer toutes les manipulations expérimentales à la suite, tandis qu'un plan expérimental emboîté fait passer les différentes manipulations à différents sujets. Le plan expérimental croisé possède une plus grande puissance statistique parce qu'il permet d'éliminer la variance due aux spécificités de chaque sujet.

la condition communicative. Ensuite, nous avons tenté de maximiser l’anonymat du punisseur dans la condition non-communicative. Les instructions exactes sont détaillées ci-dessous²²⁵.

Participants

Nous avons recruté 232 participants à partir du *Penn Subject Pool*, un système de recrutement d’étudiants de licence en psychologie à l’université de Pennsylvanie. 60 participants se sont identifiés comme des hommes et 172 comme des femmes. La moyenne d’âge était de 19,8 ans. Après que 5 étudiants ont été exclus pour avoir raté des contrôles d’attention (voir plus bas), nous nous sommes retrouvés avec 227 participants.

Méthode

Nous avons repris la méthode utilisée par Nadelhoffer et ses collègues. Comme Nadelhoffer et ses collègues, nous avons indiqué aux participants qu’ils devaient s’imaginer comme s’ils participaient à un jeu économique avec trois participants. Un premier participant devait recevoir deux dollars ; un deuxième participant avait la possibilité d’enlever ces deux dollars, sans lui-même recevoir aucun bénéfice de cette action. Nous avons demandé à nos participants de s’imaginer qu’ils étaient le troisième participant, qui avait la possibilité d’enlever à son tour de l’argent au deuxième participant si celui-ci avait choisi d’enlever de l’argent au premier participant.

Nos participants étaient successivement assignés à deux conditions, *Communication* et *Pas de communication*. Ces deux conditions étaient très proches des conditions *Ignorance* et *Connaissance* de Nadelhoffer. L’ordre dans lequel chaque participant passait chacune des deux conditions avait été randomisé. Dans la condition *Communication*, le principal changement vis-à-vis de la condition utilisée par Nadelhoffer consistait dans la possibilité d’envoyer un

²²⁵ Les trois expériences de ce chapitre ont été pré-enregistrées sur l’*Open Science Framework*. Le pré-enregistrement correspond au fait de spécifier à l’avance les analyses qui vont être conduites, afin de contrôler la flexibilité dans l’analyse des résultats. Le lien vers le pré-enregistrement des trois expériences est : https://osf.io/r8dt3/?view_only=c84b346703da4701b6e9194c952de640.

message, afin de clarifier la dimension communicative de la punition. Les participants recevaient le texte suivant :

« On vous a alloué deux dollars. Si vous choisissez de réduire de 1 dollar votre allocation, le décideur [nom donné à la personne qui a enlevé deux dollars à la première personne] ne recevra pas d'argent, et vous ne recevrez qu'un dollar. Dans ce cas, le décideur recevra aussi le message suivant : « vous ne recevrez pas le bonus de deux dollars, parce que vous avez choisi de déduire un dollar du receveur [nom donné à la première personne]. Si vous choisissez de ne pas payer un dollar, le décideur recevra alors deux dollars, et vous garderez deux dollars. Le décideur sait que le montant qu'il recevra dépend de votre choix (même s'il ne connaît pas votre identité). »

Dans la condition non-communicative, nous avons tenté de clarifier le fait que la personne potentiellement punie n'avait aucun moyen de savoir que son sort dépendait d'une troisième personne. En effet, dans le plan expérimental dont s'est servi Nadelhoffer, si le participant choisissait de ne pas punir, alors le deuxième joueur recevait un bonus de quatre dollars inattendu. Nous avons pensé qu'une telle récompense après un acte anti-social pouvait être interprété sous un mode communicatif comme une marque d'approbation, ce qui aurait pu être considéré comme une situation particulièrement négative de la part du participant. Nous avons donc choisi d'insister sur l'aspect complètement anonyme de la punition, en indiquant aux participants que le deuxième joueur savait qu'il pourrait peut-être recevoir un bonus, mais qu'il croyait que ce bonus était choisi par l'ordinateur. Les instructions exactes étaient les suivantes :

« Cependant, le décideur ne sait pas que le montant qu'il recevra dépend de votre choix. Nous avons simplement informé le décideur qu'il recevra un montant aléatoire à la fin de l'expérience, et qu'il a 50% de chances de recevoir ou bien zéro dollar, ou bien deux dollars. Bien sûr, sans qu'il le sache, le montant qu'il recevra ne sera pas aléatoire, mais dépendra de votre choix. »

Dans les deux conditions, la question principale posée aux participants était : « Que feriez-vous ? », avec deux options :

« Je choiserais de payer un dollar pour retirer deux dollars (« take \$2 away ») au décideur. »

« Je choiserais de ne pas payer un dollar. »

A des fins exploratoires, nous avons ensuite demandé aux participants quel était, selon eux, le but principal de la punition, et nous leur avons posé quelques questions démographiques (âge et genre) et une question sur leur orientation politique. Afin d'exclure les participants qui ne prêtaient pas attention au questionnaire, une question demandait simplement aux participants

de cocher la case « Agree ». Les 5 participants ayant coché une autre case ont été exclus. Aucune autre mesure n'a été collectée.

Résultats

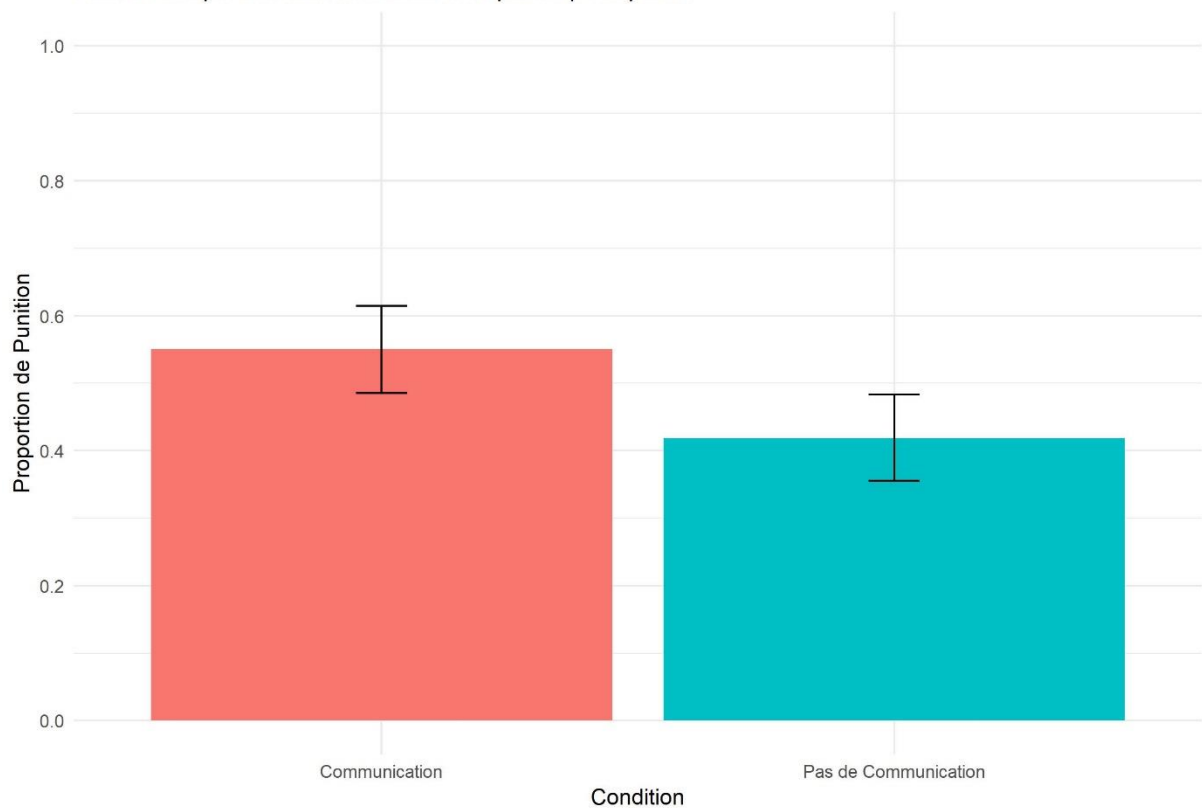
Étant donné que nous avons fait usage d'un plan expérimental croisé (*within-subject design*), cela nous permet de faire deux types différents d'analyse des données. Tout d'abord, en tirant pleinement parti du fait que les participants ont complété les deux types de condition, nous pouvons concentrer notre analyse sur les participants qui ont donné une réponse différente aux deux conditions. En nous concentrant sur ces participants-là, si nous découvrons qu'il existe plus de participants qui ont choisi de punir dans la condition *Communication*, mais pas dans la condition *Pas de communication*, que de participants qui ont fait le choix inverse, alors cela voudrait dire que la possibilité de communiquer influence la motivation punitive des participants. C'est cette manière d'analyser les données que nous avons pré-enregistré avant de recruter les participants.

Cette manière de procéder n'est cependant pas la seule possible. Il est également envisageable d'analyser cette expérience à la manière d'un plan expérimental emboîté (*between-subject design*), en analysant uniquement les réponses à la première condition vue par les participants. Cette analyse, qui correspond au choix qu'avait fait Nadelhoffer (puisque ce dernier s'était servi d'un plan emboîté), permet d'observer les réponses des participants dans un contexte où le but de l'expérience n'avait pas encore été rendu saillant par la possibilité de comparer les instructions des deux expériences. Il s'agit en quelque sorte d'une analyse des réponses « naïves », et non réflexives, des participants.

Commençons par l'analyse des résultats telle que nous l'avons pré-enregistrée, faite en comparant le nombre de participants ayant choisi de punir seulement dans la condition *Communication*, au nombre de participants ayant choisi de punir uniquement dans la condition *Pas de communication*. 52 participants ont puni dans la condition *Communication*, tout en ne punissant pas dans la condition *Pas de communication* ; 22 participants ont fait le choix inverse. La différence est hautement significative : $\chi^2(df = 1) = 12,16, p < ,001$. En termes de proportions, cela correspond au fait que 55% des participants ont choisi de punir dans la condition *Communication*, contre 42% dans la condition *Pas de Communication* (Figure 8). Cette augmentation de 31% du taux de punition n'est pas négligeable, même s'il est clair que

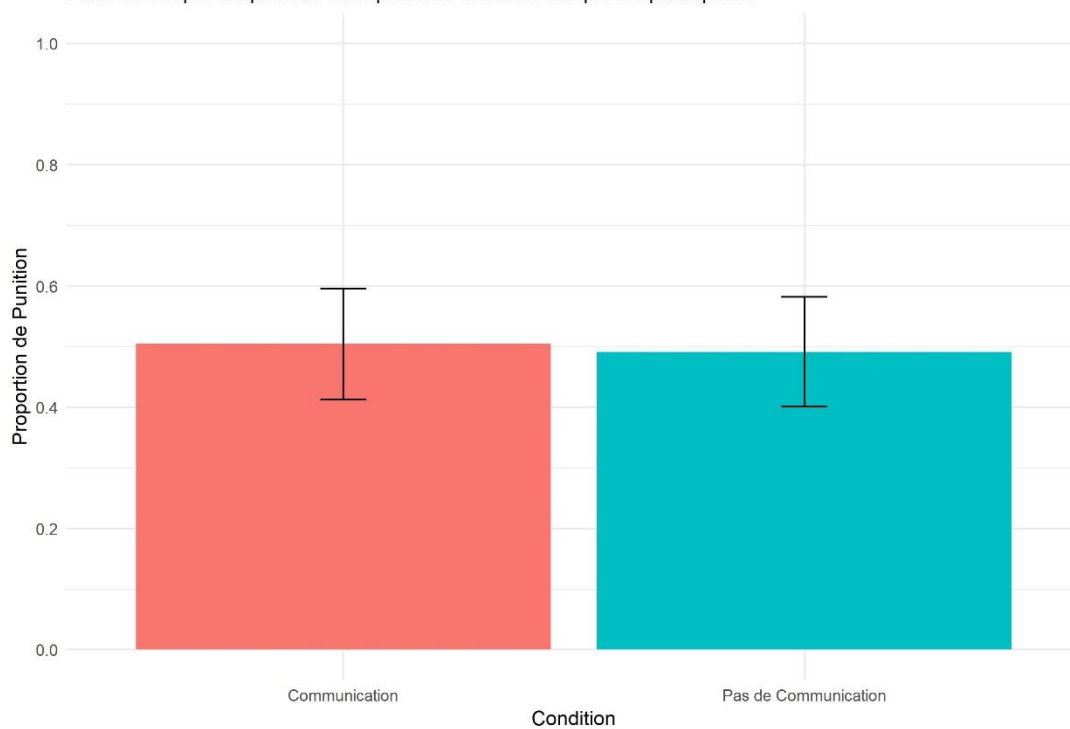
la majorité des participants ont fait leur choix pour des raisons autres que des objectifs communicatifs. Si l'on se concentre sur toutes les réponses que les participants ont données, il est donc clair que la possibilité de communiquer augmente la motivation à punir.

Fig. 8. Punition dans les deux conditions de l'expérience 1.
Prise en compte des deux conditions vues par les participants.



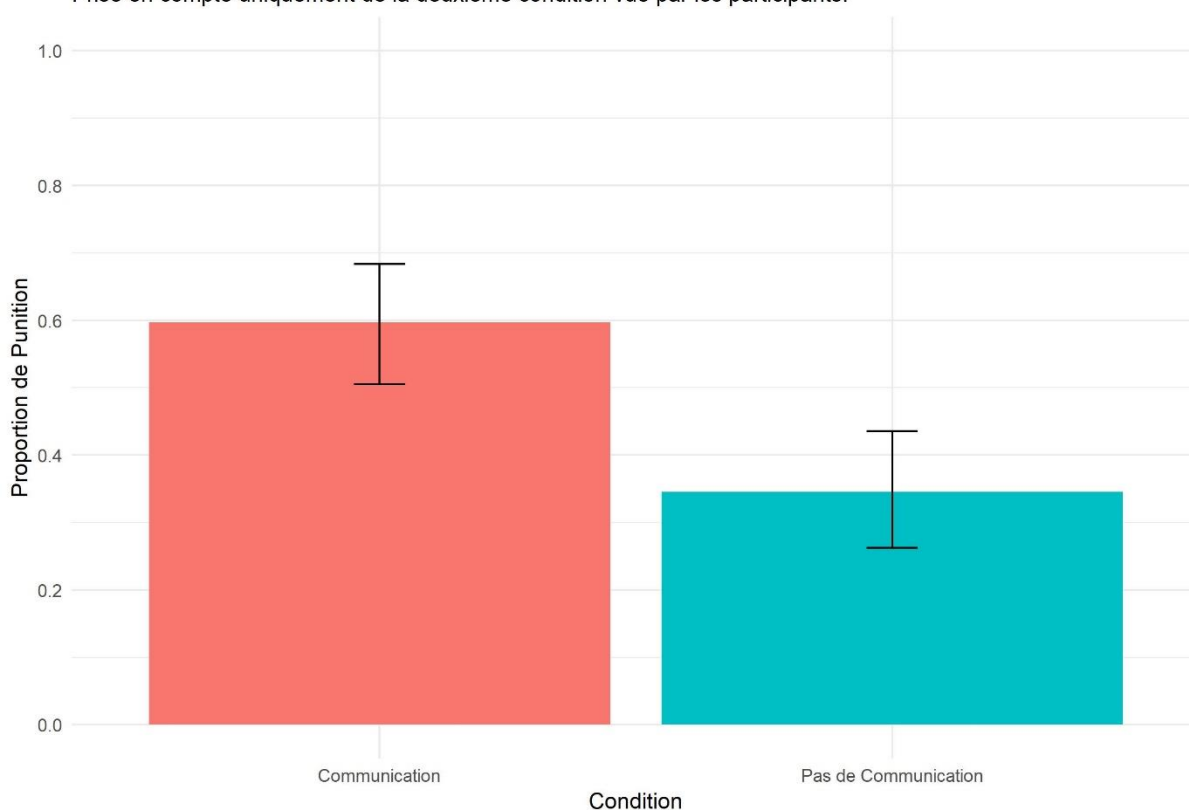
La situation est cependant différente si l'on considère uniquement les réponses à la première condition que les participants ont vue. Dans cette situation, la différence entre les deux conditions est négligeable : 50% des participants ont choisi de punir dans la condition *Communication*, contre 49% dans la condition *Pas de Communication* (voir Figure 9). Une telle différence n'est évidemment pas statistiquement significative. On retrouve ici exactement les résultats de Nadelhoffer : dans les deux cas, la possibilité de communiquer n'a pas eu d'impact sur la motivation à punir tant que le rôle de la communication n'a pas été particulièrement mis en avant.

Fig. 9. Punition dans les deux conditions de l'expérience 1.
Prise en compte uniquement de la première condition vue par les participants.



La différence entre les résultats globaux et les résultats obtenus en se concentrant sur la première condition tient bien sûr au fait que les participants n'ont pas réagi du tout de la même manière à la deuxième condition qu'ils ont vue. Les participants ayant d'abord vu la condition *Pas de Communication* ont choisi de punir à 60% dans la deuxième condition, à savoir la condition *Communication*. A l'inverse, les participants qui ont d'abord vu la condition *Communication* ont puni à 35% dans la condition *Pas de Communication* (voir figure 10). Alors qu'il existait un écart négligeable de 1 point de pourcentage quand on compare les premières conditions que les participants ont vu, cet écart passe à 25 points de pourcentage quand on considère les deuxièmes conditions vues par les participants. La saillance de la manipulation expérimentale a eu des conséquences extrêmement fortes sur le comportement des participants.

Fig. 10. Punition dans les deux conditions de l'expérience 1.
Prise en compte uniquement de la deuxième condition vue par les participants.



Discussion

Cette première expérience offre des résultats contrastés sur l'effet de la possibilité de communiquer sur la motivation à punir. Dans la limite du paradigme expérimental des jeux économiques, il semble que la possibilité de communiquer n'augmente la motivation à punir que si l'importance de la communication est rendue saillante, et si le participant a la possibilité de comparer deux situations de punition, l'une où la communication est possible, l'autre où la communication est impossible.

Des arguments opposés peuvent être formulés pour savoir quel type de résultat doit être considéré comme étant le plus valide d'un point de vue écologique. En défense du plan expérimental emboîté (*between-subject design*), dans lequel nous n'avons pas constaté de différence entre les deux conditions, on peut mettre en avant le fait que les réponses des participants sont davantage spontanées s'ils n'ont pas la possibilité de comparer les deux expériences. Une critique fréquente des plans expérimentaux croisés (*within-subject designs*)

est que ceux-ci créent un biais expérimental, dit biais de désirabilité sociale, où les participants seraient conduits à donner la réponse qu'ils pensent être celle que l'expérimentateur désire. Si l'on est convaincu par ces arguments, on pourra penser que la possibilité de communiquer n'a en général pas un impact très fort sur la motivation à punir.

A l'inverse, en défense du plan expérimental croisé (*within-subject design*), qui a conduit à une très nette différence entre les deux conditions, on peut mettre en avant que, dans la plupart des cas de punition dite de la troisième partie, les individus ont la possibilité de réfléchir aux conditions dans lesquelles ils vont punir quelqu'un. Si un patron veut punir un de ses salariés, il peut comparer les situations dans lesquelles il pourrait le faire de façon anonyme, et les situations où il pourrait le faire de façon explicite. La même chose peut être dite des situations où un enseignant veut punir un élève, ou un policier un délinquant. Dans ce cadre, il serait possible de dire que la condition la plus écologiquement valide est celle où les individus ont la possibilité de comparer différentes situations, et où ils ont le loisir de choisir la situation où ils peuvent communiquer à l'individu puni. Si l'on suit ce raisonnement, la possibilité de communiquer aurait bien en général un impact sur la motivation de punir.

Je ne prétends pas trancher ces questions ici. Je vais maintenant passer à la présentation des deux autres expériences qui visent à estimer l'impact de la communication sur la légitimité et sur la conception de la punition.

Expérience 2

Nous venons donc de voir que la possibilité de communiquer a un impact ambigu sur la volonté de punir. Cette question est cependant indépendante de celle portant sur la perception de la justice d'une punition avec ou sans communication. De même, cette question n'a pas de lien avec la question de savoir si la punition doit faire référence ou non à la communication, c'est-à-dire, si la communication fait partie du concept de punition.

En ce qui concerne le concept de punition, nous pouvons partir d'une définition classique, et fréquemment reprise, proposée par Hart dans son *Punishment and Responsibility* :

« (i) [La punition] doit inclure de la souffrance ou d'autres conséquences normalement considérées désagréables.

(ii) Elle doit être causée par une offense à l'encontre de règles légales.

(iii) Elle doit être menée à l'encontre d'un délinquant réel ou supposé, et effectuée en raison de son offense.

(iv) Elle doit être administrée de façon intentionnelle par des êtres humains autres que le délinquant.

(v) Elle doit être infligée et administrée par une autorité constituée par un système légal contre lequel l'offense a été commise. »

Les règles (i), (iii) et (iv) me semblent incontournables pour définir la punition. Les règles (ii) et (v) me semblent davantage contestables, et révèlent le fait que Hart a essentiellement cherché à constituer une théorie de la punition légale, et non pas de la punition en général. Ces deux conditions incluent deux aspects principaux : le fait que le délinquant doit avoir effectué un comportement contre la loi, et le fait que le délinquant soit puni par une autorité légalement constituée. Que l'idée de punition implique l'existence d'une autorité possède une certaine plausibilité ; par contre, il me semble invraisemblable que l'idée de punition requière l'existence d'une loi. Si un parent prive de dessert son enfant parce que cet enfant s'est mal comporté, il me semble assez clair que le concept de punition s'applique ici, alors même que les caprices enfantins ne sont que rarement interdits par le code juridique d'un État. Je reprendrai donc essentiellement ici les critères de Hart, sans reprendre l'exigence de légalité qu'il a défendue.

En ce qui nous concerne, il est important de remarquer que les critères utilisés par Hart n'incluent nulle part l'idée que la punition doit avoir un objectif communicatif, ou l'idée que la personne punie doit avoir conscience qu'elle a été punie. La punition, selon ces critères, doit être infligée intentionnellement, mais cela n'implique en rien que la personne punie prenne conscience qu'elle a été punie. Le but des deux prochaines expériences est donc en partie d'examiner dans quelle mesure les critères de Hart doivent être modifiés pour inclure une dimension communicative.

Pour étudier l'impact de la communication sur la légitimité et la conceptualisation de la punition, nous avons présenté aux participants divers cas d'actions négatives effectuées par plusieurs acteurs. Ces actions négatives étaient toujours suivies d'une sanction. Pour étudier l'impact de la communication, nous avons varié dans quelle mesure la personne sanctionnée avait conscience des buts et des raisons sous-tendant la sanction.

Participants

Nous avons recruté 201 participants à partir de la *Penn Subject Pool*, un système de recrutement d'étudiants de licence en psychologie à l'université de Pennsylvanie. 74 participants se sont identifiés comme des hommes et 127 comme des femmes. La moyenne d'âge était de 19,91 ans. Après que 6 étudiants ont été exclus pour avoir raté des contrôles d'attention (voir plus bas), nous nous sommes retrouvés avec 195 participants.

Méthode

Dans cette expérience, nous avons utilisé un plan expérimental croisé, où chaque participant répondait à tour de rôle aux six différentes conditions. Les conditions variaient par le type d'histoire (trois types d'histoire différents) et par le degré de communication (*Communication* versus *Pas de communication*). Dans chaque condition, un individu effectuait une action négative, avant d'être sanctionné par une tierce personne. Dans la condition *Communication*, la personne qui a choisi de sanctionner le comportement donnait clairement les raisons de la sanction. Dans la condition *Pas de communication*, aucune raison n'était donnée, et la personne sanctionnée ne comprenait pas pourquoi elle avait été punie.

Les trois histoires portaient sur un employé sanctionné par son patron, sur un étudiant sanctionné par son professeur, et sur un jeune homme dans un parc sanctionné par un passant. L'employé a saboté le travail d'un de ses collègues, et a perdu un bonus de ce fait. L'étudiant a triché à un examen, et a perdu des points à cause de cela. Le jeune homme a volé un téléphone portable, et le passant a jeté à la poubelle le portefeuille de ce jeune homme en profitant d'un moment d'absence de celui-ci. Ces trois histoires ont été utilisées pour diversifier les situations et pour éviter que les réponses soient liées à un cas en particulier. Nous avons évité des cas de sanction légale par peur que les réponses des participants soient motivées uniquement par des soucis de conformité au cadre juridique, et qu'elles ne reflètent pas un concept général de punition. Le cas du jeune homme a été ajouté pour fournir un exemple où la personne qui sanctionne n'était pas une personne en position d'autorité, afin d'examiner si l'impact de la communication était similaire dans les deux cas.

Pour comparer les conditions *Communication* et *Pas de communication*, je traduis ici la fin de la vignette portant sur le patron sanctionnant son employé :

« A l'insu de Bill, le patron de Bill apprend qu'il a saboté le travail de son collègue, et, en conséquence, décide de ne pas donner de bonus à Bill cette année-là.

[Communication] Après avoir convoqué Bill, le patron de Bill lui dit qu'il n'a pas eu le bonus parce qu'il a saboté le travail de son collègue.

[Pas de communication] Cependant, le patron de Bill ne lui a pas dit les raisons pour lesquelles il n'a pas reçu de bonus, et John n'a donc jamais compris la vraie raison pour laquelle il n'a pas reçu de bonus. »

Après avoir lu chacune des histoires, les participants ont répondu à une série de questions, dans lesquelles ils devaient indiquer dans quelle mesure ils étaient d'accord avec une série de phrases. L'accord était mesuré sur une échelle de 1 à 7, 1 indiquant « Pas du tout d'accord », et 7 « Parfaitement d'accord ».

Les questions étaient les suivantes²²⁶ :

« Bill a été puni pour avoir saboté le travail de son collègue.

Bill a été traité de façon juste par son patron.

Je suis parfaitement satisfait par la manière dont le patron de Bill a géré la situation.

Bill a reçu ce qu'il méritait.

Bill méritait un traitement plus dur.

Bill méritait un traitement plus doux. »

Les trois premières phrases sont les plus utiles pour notre propos. La première renvoie à la question conceptuelle : si les participants indiquent un plus fort accord avec cette phrase dans la condition *Communication* que dans la condition *Pas de communication*, il s'agirait d'une indication que la communication fait partie du concept de punition. La deuxième phrase renvoie à la légitimité, et la troisième à la satisfaction vis-à-vis du traitement de l'employé.

A la suite de ces questions, nous avons posé quelques questions démographiques (âge et genre) et une question sur leur orientation politique. Afin d'exclure les participants qui ne prêtaient pas attention au questionnaire, une question demandait simplement aux participants de cocher la case « Agree ». Les 6 participants ayant coché une autre case ont été exclus. Aucune autre mesure n'a été collectée.

²²⁶ Le texte des phrases était à chaque fois adapté au contexte des vignettes. Le lecteur peut voir ci-dessus les questions telles qu'elles ont été posées dans le cas de la vignette portant sur le contexte de l'entreprise.

Résultats

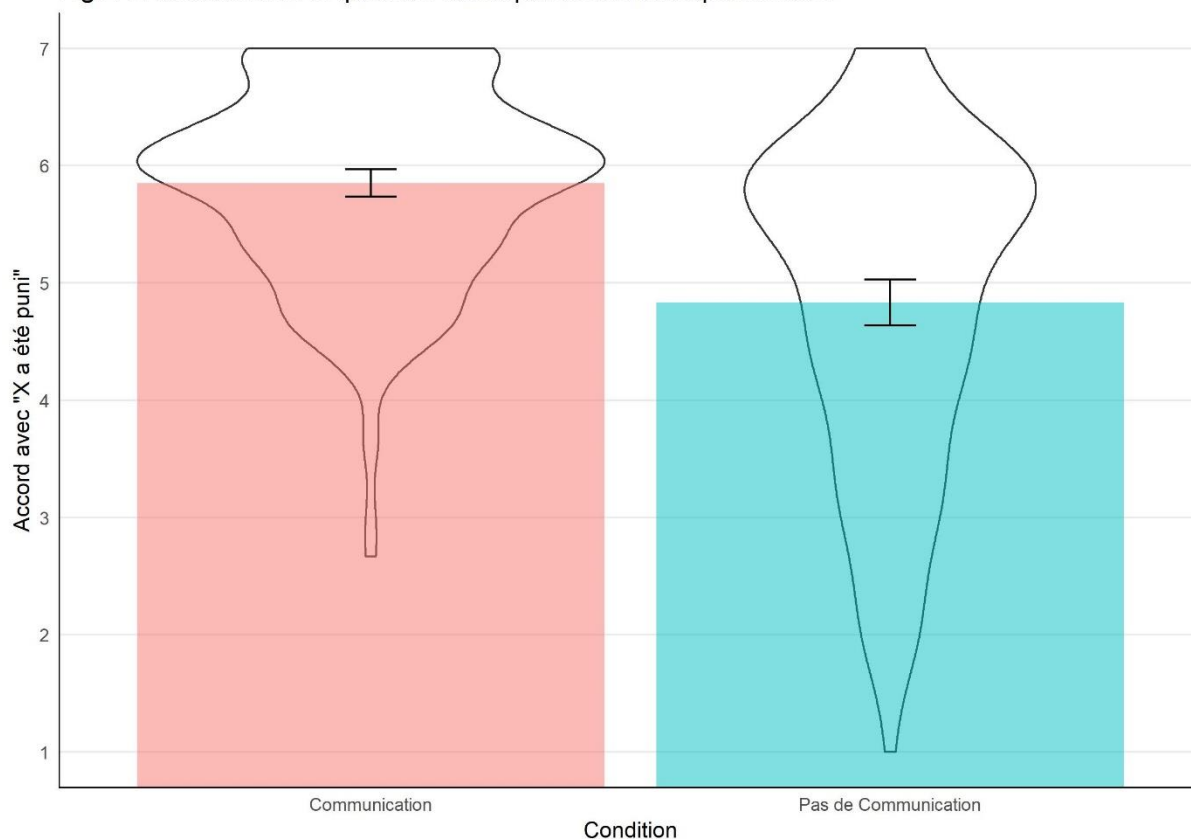
Comme dans le cas de la première expérience, il est possible d'analyser ces résultats ou bien en prenant pleinement en compte le fait qu'il s'agit d'un plan expérimental croisé (*within-subject design*) en combinant les premières et deuxièmes réponses des sujets, ou comme un plan expérimental emboîté (*between-subject design*), en se concentrant sur les réponses faites par les sujets au premier scénario qu'ils ont vu. Dans ce cas précis, les deux types d'analyse donnent globalement des réponses concordantes. Quel que soit le mode d'analyse, dans le cas des trois questions-clés, les participants ont bien davantage tendance à décrire une sanction comme étant une punition si elle implique une communication (question conceptuelle) ; ils ont également davantage tendance à décrire la situation comme étant juste (légitimité) ; ils se disent également davantage satisfaits par le comportement de la personne qui sanctionne. Le seul cas où les analyses croisée et emboîtée donnent des résultats différents concerne le cas du passant punissant le jeune dans un parc, comme nous allons le voir plus bas ; dans ce cas, les différences sont très nettes dans le cas où les participants ont eu accès aux deux scénarios, mais beaucoup moins nettes dans le cas où ils ont eu seulement accès au premier scénario, probablement parce que la punition causée par quelqu'un qui ne dispose d'aucune autorité sur le sujet n'était pas perçue comme très légitime de toute façon.

Je vais commencer par présenter les résultats à la question conceptuelle. Si l'on combine les trois expériences en une, on obtient un accord moyen avec la phrase « Bill a été puni » de 5,85 dans le cas où il y a eu communication, et un accord de 4,83 dans le cas où il n'y a pas eu communication (Voir Figure 11). Il s'agit donc d'une différence extrêmement forte. Cette différence est significative dans chacune des trois vignettes employées, et est robuste à une analyse faite à partir de l'ensemble des résultats, comme à partir seulement des premiers scénarios vus par les participants²²⁷. Je reviendrai dans la discussion sur l'analyse de ces résultats.

²²⁷ Commençons par les t-tests pour échantillons appariés de Welch, utilisés pour comparer les différentes conditions que les participants ont vues. Dans le cas de l'entreprise, l'accord était beaucoup plus fort dans la condition *Communication* : $\Delta M = 0,95$, 95% CI [0,72, 1,18], $t(194)=8,10$, $p<,001$. Un résultat similaire se retrouve dans la vignette de l'enseignant : $\Delta M = 1,29$, 95% CI [1,02, 1,55], $t(194)=9,49$, $p<,001$, et dans le cas de la vignette du passant : $\Delta M = 0,82$, 95% CI [0,55, 1,08], $t(194)=6,16$, $p<,001$.

Passons maintenant aux t-tests pour échantillons indépendants de Welch, utilisés pour comparer la première vignette vue par les participants. Dans le cas de l'entreprise, on a là aussi un accord plus fort dans le cadre de la vignette *Communication* : $\Delta M=0,73$, 95% CI [0,32, 1,15], $t(191,68)=3,49$, $p=,001$. Cette différence se retrouve aussi dans le cas de l'enseignant : $\Delta M=1,22$, 95% CI [0,81, 1,64], $t(179,31)=5,81$, $p<,001$, et dans le cas du passant : $\Delta M=0,67$, 95% CI [0,20, 1,15], $t(187,98)=2,79$, $p=,006$.

Fig. 11. Résultats de la question conceptuelle dans l'expérience 2.

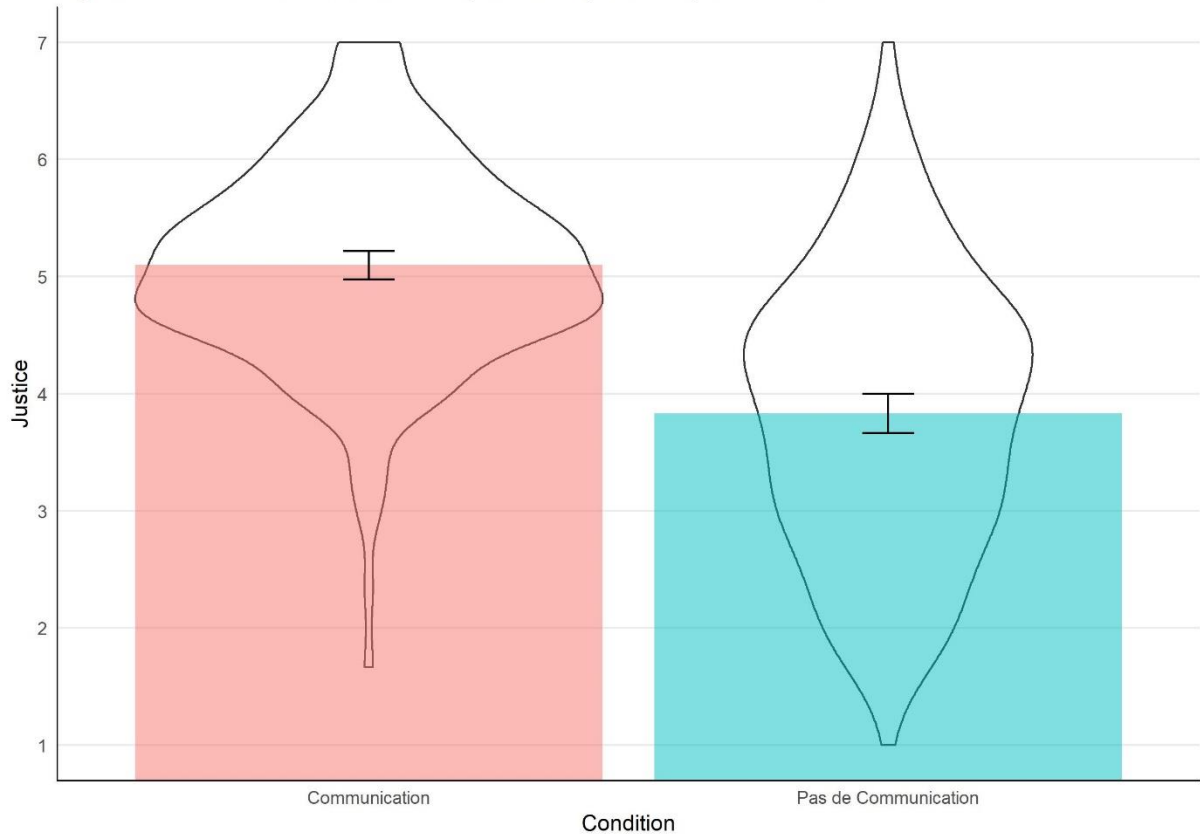


Passons maintenant aux résultats à la question de la légitimité. Si l'on combine les trois expériences en une, on obtient un accord moyen avec la phrase « Bill a été traité de façon juste » de 5,10 dans le cas où il y a eu communication, et un accord de 3,83 dans le cas où il n'y a pas eu communication (Voir Figure 12). Il s'agit donc ici aussi d'une différence extrêmement forte. Cette différence est significative dans chacune des trois vignettes employées dans le cas de l'étude faite à partir de l'échantillon apparié ; elle est robuste dans le cas de l'entreprise et de l'enseignant même si l'on se concentre uniquement sur les premiers scénarios vus. Dans le cas du passant, il n'y a pas de différence significative si l'on se concentre uniquement sur le premier scénario vu par les participants, probablement parce que les individus considèrent qu'il s'agit d'une sanction très injuste dans tous les cas, étant donné que le passant n'avait pas d'autorité réelle sur la question²²⁸.

²²⁸ Dans le cas des t-tests pour échantillons appariés de Welch, dans la vignette *Entreprise*, l'accord est beaucoup plus grand dans la condition *Communication* : $\Delta M=1,49$, 95% CI [1,25, 1,73], $t(194)=12,28$, $p<,001$. Ce résultat se retrouve dans le cas de l'enseignant : $\Delta M=2,01$, 95% CI [1,75, 2,27], $t(194)=15,43$, $p<,001$, et dans le cas du passant : $\Delta M=0,29$, 95% CI [0,09, 0,50], $t(194)=2,79$, $p=,006$.

En ce qui concerne le test fait sur les échantillons indépendants, on a là aussi un accord plus élevé dans le cas de l'entreprise : $\Delta M=0,88$, 95% CI [0,51, 1,25], $t(185,27)=4,67$, $p<,001$. Ce résultat se retrouve dans le cas de

Fig. 12. Justice du traitement de la personne punie, expérience 2.

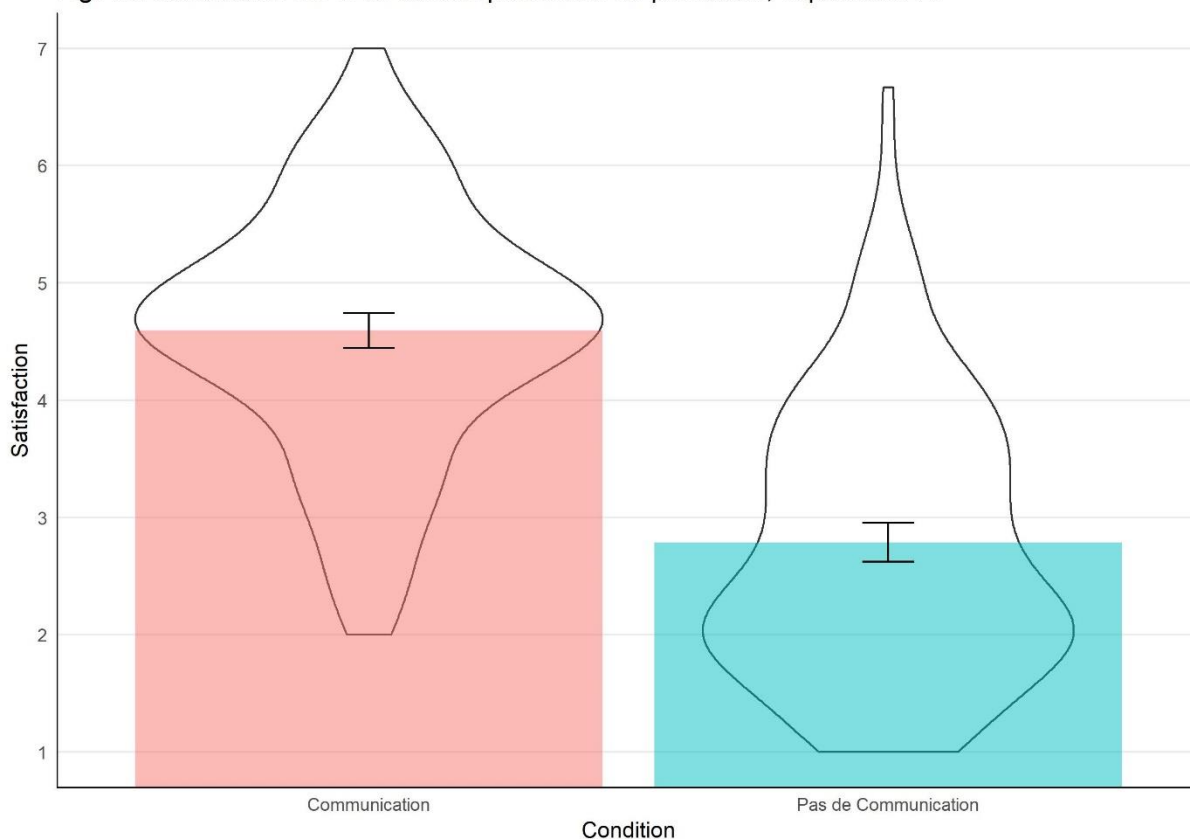


Les résultats portant sur la question de la satisfaction donnent des résultats très proches. Si l'on combine les trois expériences en une, on obtient un accord moyen avec la phrase « Je suis parfaitement satisfait par la manière dont [l'acteur] a géré la situation » de 4,59 dans le cas où il y a eu communication, et un accord de 2,79 dans le cas où il n'y a pas eu communication (Voir Figure 13). Il s'agit donc ici encore d'une différence très forte. Cette différence est significative dans chacune des trois vignettes employées dans le cas de l'étude faite à partir de l'échantillon apparié ; elle est robuste dans le cas de l'entreprise et de l'enseignant même si l'on se concentre uniquement sur les premiers scénarios vus. Dans le cas du passant, il n'y a pas de différence significative si l'on se concentre uniquement sur le premier scénario vu par les participants, probablement parce que les individus considèrent qu'il s'agit d'une sanction très injuste dans tous les cas, étant donné que le passant n'avait pas d'autorité réelle sur la question²²⁹.

l'enseignant : $\Delta M=1,51$, 95% CI [1,09, 1,94], $t(184,07)=6,98$, $p<,001$. Par contre, il n'existe pas de différence significative dans le cas du passant : $\Delta M=-0,13$, 95% CI [-0,55, 0,29], $t(192,91)=-0,62$, $p=,533$.

²²⁹ Dans le cas des t-tests pour échantillons appariés de Welch, dans la condition *Entreprise*, on a un accord beaucoup plus élevé dans la condition *Communication* : $\Delta M=2,20$, 95% CI [1,93, 2,47], $t(194)=16,35$, $p<,001$. Ce résultat se retrouve dans le cas de l'enseignant : $\Delta M=2,69$, 95% CI [2,43, 2,96], $t(194)=20,21$, $p<,001$, et dans le cas du passant : $\Delta M=0,53$, 95% CI [0,31, 0,74], $t(194)=4,84$, $p<,001$.

Fig. 13. Satisfaction vis-à-vis du comportement du punisseur, expérience 2.



Discussion

Les résultats de l'expérience portent sur deux sujets : l'influence de la communication sur la perception d'une punition légitime, d'une part, et l'influence de la communication sur le concept de punition. Je pense que le traitement de la légitimité est plus clair que celui de la question conceptuelle, et je commencerai donc par ce premier point.

En ce qui concerne la question de la légitimité, le résultat des expériences est très net. Une punition a été perçue comme étant beaucoup plus juste si elle a été effectuée de façon communicative que si elle a été infligée sans communication. La différence moyenne de 1,3 sur une échelle de 1 à 7 est très forte. Les participants ont été également beaucoup plus satisfaits

Pour les échantillons indépendants, on obtient des résultats similaires dans les vignettes de l'entreprise et de l'enseignant. Dans le cas de l'entreprise, l'accord est aussi plus élevé dans la condition *Communication* : $\Delta M=1,39$, 95% CI [0,93, 1,85], $t(192,87)=5,96$, $p<,001$. Ce résultat est identique dans le cas de l'enseignant : $\Delta M=2,09$, 95% CI [1,65, 2,53], $t(193,00)=9,38$, $p<,001$. Là encore, pour le passant, le résultat est le seul à ne pas être significatif : $\Delta M=-0,16$, 95% CI [-0,56, 0,24], $t(192,93)=-0,79$, $p=,431$.

par l'action de la personne sanctionnant dans le cas de la condition communicative. De tels résultats apportent un soutien net aux théories communicatives de la peine.

Bien sûr, il est possible de fournir une explication alternative de ces résultats en termes de théories dissuasives de la peine. S'il existe une communication, il est évident que la personne punie aura probablement moins de chances à l'avenir de recommencer des actes ainsi réprimandés. Nos résultats par eux-mêmes ne permettent pas de trancher entre une explication communicationnelle et dissuasive. Cependant, comme nous l'avons vu, il existe maintenant une littérature assez importante sur le fait que les participants aux expériences sont en général assez peu sensibles aux facteurs dissuasifs dans leur estimation de la justice d'une peine, et sont beaucoup plus sensibles à des facteurs apparentés au rétributivisme. En ce sens, une interprétation communicative de nos résultats semble plus probable qu'une explication en termes dissuasifs.

L'interprétation de nos résultats en ce qui concerne la question conceptuelle est plus délicate. Bien sûr, comme dans le cas de la question de la légitimité, nous avons obtenu des scores bien plus élevés d'accord avec la phrase « Bill a été puni » dans le cas des conditions *Communication* que dans le cas des conditions *Pas de communication*. Cependant, il faut bien remarquer que les moyennes restent hautes dans les deux cas. La moyenne est de 5,85 dans le cas où il y a eu communication, et un accord de 4,83 dans le cas où il n'y a pas eu de communication. Cette dernière moyenne de 4,83 cache de profondes hétérogénéités. Dans le cas des conditions *Communication*, il existe un consensus sur le fait qu'il y a bien eu punition ; 87% des participants ont indiqué un accord moyen supérieur ou égal à 5, et seulement 1% des participants ont donné une réponse moyenne inférieure ou égale à 3. A l'inverse, il existe beaucoup plus d'étalement des réponses dans le cas des conditions *Pas de communication*. On trouve à la fois des participants ayant indiqué qu'il s'agit clairement d'une punition, et des participants ayant indiqué qu'il ne s'agit clairement pas d'une punition. Dans le cas *Pas de communication*, 15% des gens ont indiqué une réponse inférieure ou égale à 3, tandis que 57% ont donné une réponse supérieure ou égale à 5.

Dans cette situation, comment interpréter ces résultats ? Une première conclusion est que seule une minorité de participants semblent considérer que la communication est une condition nécessaire de l'application du concept de punition. Pour la majorité, le concept semble pouvoir s'appliquer même dans des situations non-communicatives. Cependant, on observe un accord moyen global plus bas avec la question conceptuelle dans le cas de la condition *Pas de communication*, et cet accord moyen plus bas n'est pas uniquement dû aux participants qui ont jugé que la communication était une condition nécessaire de l'emploi du terme de punition. En

effet, la grosse majorité des participants (72%) a donné une réponse moyenne supérieure aux cas *Communication* qu'aux cas *Pas de communication*. Dans ce contexte, il semble bien que, même si la communication n'est pas une condition nécessaire pour l'application du terme de punition, il s'agit quand même d'une part cruciale du cas prototypique de l'idée de punition.

On pourrait cependant objecter que l'expérience précédente mélange deux aspects assez différents sous le terme général de « communication ». Pour reprendre l'exemple de l'entreprise, il y a deux choses qui diffèrent assez nettement entre les conditions *Communication* et *Pas de communication*. La condition *Communication* combine à la fois un aspect d'information pure (l'employé apprend pourquoi il a été sanctionné), et un aspect de justice procédurale (l'information a été transmise de la bonne manière par l'autorité compétente). Ces deux aspects sont tous les deux pertinents pour déterminer la légitimité d'une punition selon la théorie communicative de la peine, mais on pourrait dire que les deux aspects ne sont pas également pertinents. L'aspect informatif est ce qui correspond au cœur d'une théorie communicative de la peine. L'aspect de légitimité procédurale est également crucial pour toute théorie communicative, mais représente un aspect de transparence dans la prise de décision qui possède probablement une signification plus générale, et qui n'est pas réductible à l'attrait des théories communicationnelles de la peine.

Dans l'expérience 3, nous tentons de distinguer ces deux aspects, en créant des conditions qui maintiennent séparés l'aspect purement informatif de l'aspect de légitimité procédurale.

Expérience 3

Notre expérience 3 a deux objectifs. Tout d'abord, il s'agit d'effectuer une réplique de nos résultats de *l'Expérience 2*, afin de vérifier la robustesse de nos résultats sur un nouvel échantillon. Ensuite, il s'agit de mieux comprendre les résultats de notre *Expérience 2*, afin de déterminer dans quelle mesure nos résultats ont été davantage dus à l'aspect informatif de la communication, ou au fait que la communication sert une fonction de justice procédurale. Pour tester ces deux hypothèses concurrentes, nous avons ajouté une troisième condition expérimentale aux conditions de *l'Expérience 2*, dans laquelle la personne punie apprend par hasard la raison pour laquelle elle a été sanctionnée.

Participants

Nous avons recruté 970 participants à partir d'*Amazon Mechanical Turk*, une plate-forme de recrutement en ligne. 539 participants se sont identifiés comme des hommes, 426 comme des femmes, et 5 ont choisi la catégorie « autre ». La moyenne d'âge était de 37,5 ans. Après avoir exclu 50 participants qui ont raté le contrôle d'attention, ont répondu trop vite (en moins de 50 secondes) ou ont fourni un mot incorrect comme seule réponse dans leurs justifications (« bp »), nous nous sommes retrouvés avec 920 participants²³⁰.

Méthode

Dans l'expérience 3, nous avons choisi de reprendre en grande partie les mêmes stimuli et la même méthode que dans l'expérience 2, avec plusieurs modifications à noter. Tout d'abord, étant donné que nous avons obtenu des résultats très similaires entre les différentes histoires de l'expérience 2, nous avons choisi de nous servir uniquement de la vignette « entreprise », où un patron sanctionne son employé. Deuxièmement, nous avons utilisé cette fois-ci un plan expérimental emboîté (*between-subjects design*) ; la perte de puissance statistique comparée à la deuxième expérience est ici compensée par l'usage d'un échantillon bien supérieur. Troisièmement, nous avons ajouté les questions suivantes à des fins d'analyse exploratoire :

« Bill a probablement appris une leçon.

Il est peu probable que Bill sabote le travail de son collègue de nouveau à l'avenir.

Bill a été traité de façon dure. »

Enfin, la modification la plus importante a consisté à ajouter une troisième condition, nommée *Communication indirecte*. Cette condition consiste à indiquer que la personne a reçu l'information, mais uniquement de manière indirecte, sans qu'il y ait eu intention de la part de la personne sanctionnant de transmettre l'information. En d'autres termes, la conséquence en termes éducatifs et communicationnels est la même (la personne a bien reçu l'information), mais l'acte de punir n'a pas respecté les mêmes contraintes procédurales. La fin de la vignette était la suivante (le nom du protagoniste était Jim dans cette condition) :

²³⁰ Ce changement de critères d'exclusion, plus strict par rapport aux expériences 1 et 2, a été pré-enregistré.

« Cependant, le patron de Jim décide de ne pas dire à Jim pourquoi il n'a pas reçu de bonus. Au début, Jim ne comprend pas pourquoi il n'a pas reçu de bonus, mais quelques jours après la décision, Jim apprend la vraie raison par hasard en interrogeant un autre collègue (qui avait surpris une conversation du patron discutant du sujet). »

Résultats

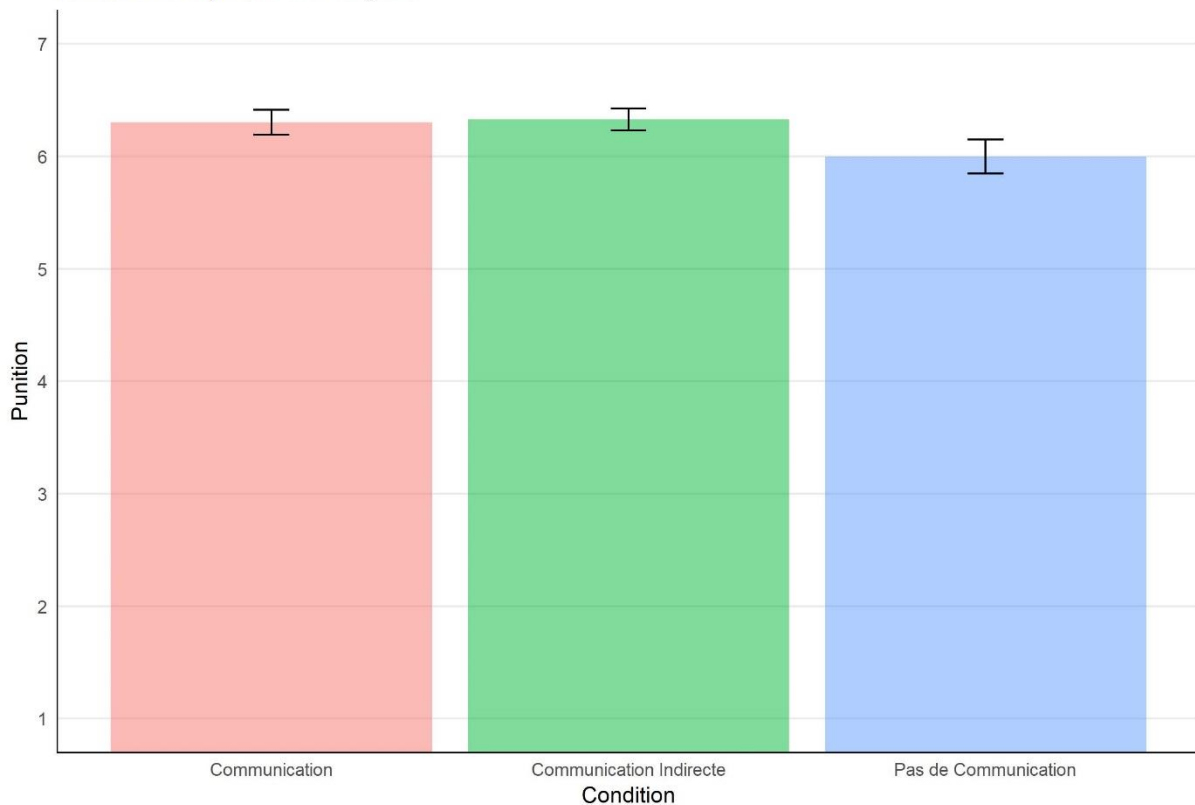
Tout d'abord, je vais commencer par présenter la comparaison entre la condition *Pas de communication* et la condition *Communication*, en laissant de côté pour l'instant la nouvelle condition. Dans ce cadre, l'expérience 3 réplique en tous points les résultats de l'expérience 2. En effet, comme dans l'expérience 2, les participants ont davantage tendance à décrire la situation comme un cas de punition s'il y a eu communication, comparée à la condition où il n'y pas eu de communication ($\Delta M = 0,31$, 95% CI [0,12, 0,49], $t(557,46) = 3,21$, $p = ,001$). La différence entre les deux conditions est cependant plus faible que dans l'expérience 2. Une explication possible est le fait que les participants ont eu beaucoup plus tendance à décrire la situation comme une situation de punition que dans l'expérience 2, provoquant un effet de plafond.

De plus, les participants se sont déclarés beaucoup plus satisfaits vis-à-vis du comportement du patron dans le cas de la condition *Communication* que dans le cas de la condition *Pas de communication* ($\Delta M = 0,66$, 95% CI [0,38, 0,94], $t(592,84) = 4,65$, $p < ,001$), et ont indiqué un plus grand accord avec l'idée que le salarié avait été traité de façon juste ($\Delta M = 0,41$, 95% CI [0,20, 0,62], $t(591,13) = 3,92$, $p < ,001$). Si l'on utilise maintenant une des nouvelles questions, portant sur la perception du succès de la communication : « Bill a probablement appris une leçon. », on obtient une très grande différence entre les deux conditions. La moyenne d'accord dans le cas de la condition *Communication*, à 4,80, tombe à 3,10 dans le cas de la condition *Pas de communication*. Cette différence est bien sûr significative ($\Delta M = 1,70$, 95% CI [1,43, 1,96], $t(602,18) = 12,65$, $p < ,001$). Même si de tels résultats ne sont pas fondamentalement très surprenants, ils indiquent bien qu'il y a la perception d'un succès de la communication.

Si l'on compare maintenant la nouvelle condition, *Communication indirecte*, aux deux autres conditions, les résultats sont, pour une bonne partie des questions, très proches de ceux obtenus dans le cas de la condition *Communication*. Pour le cas de la question conceptuelle, interrogeant

sur l'accord avec la description de la situation comme une punition, il n'existe aucune différence significative entre la condition *Communication* et la condition *Communication indirecte*, et l'intervalle de confiance à 95% n'inclut que de très faibles différences entre les deux conditions ($\Delta M = -0,02$, 95% CI $[-0,17, 0,12]$, $t(605,40) = -0,33$, $p = ,742$). A l'inverse, il existe une différence significative entre la condition *Communication indirecte* et la condition *Pas de communication*, la condition *Communication indirecte* étant davantage perçue comme une punition ($\Delta M = 0,33$, 95% CI $[0,15, 0,51]$, $t(519,61) = 3,61$, $p < ,001$). Ces résultats indiquent que ce qui est crucial pour décrire une situation comme punition est le fait que la personne punie comprenne les raisons de la punition ; le fait que la communication provienne d'une source légitime semble avoir moins d'importance. Les différentes moyennes sont présentées dans la figure 14.

Fig. 14. Accord avec la question conceptuelle, expérience 3.
Accord avec la phrase "X a été puni"

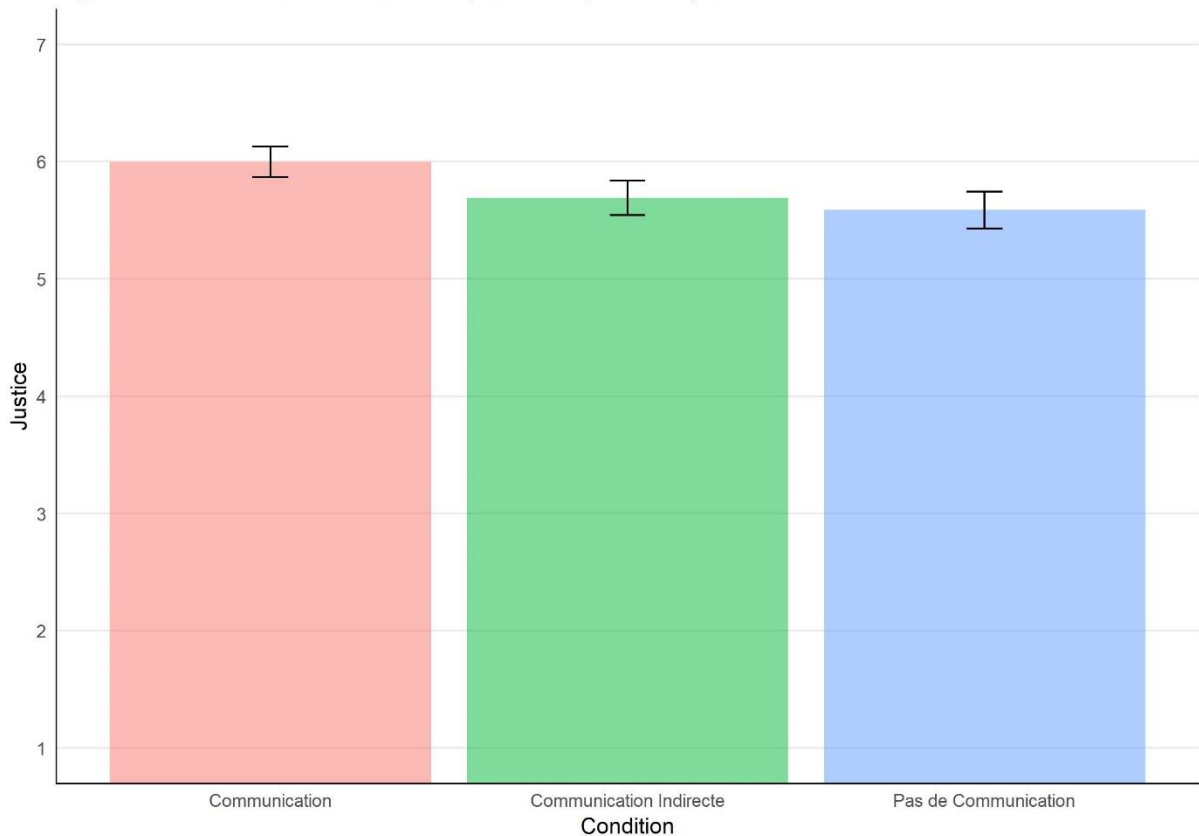


De la même manière, les réponses pour la question portant sur l'impact éducatif de la peine : « Bill a probablement appris une leçon. », donnent des résultats très similaires dans le cas de la condition *Communication* et dans le cas de la condition *Communication Indirecte*. Là aussi, il n'existe pas de différence significative (la moyenne passant de 4,80 à 4,71), et l'intervalle de confiance à 95% n'inclut pas de différence très forte entre les deux conditions ($\Delta M = 0,09$, 95% CI $[-0,16, 0,33]$, $t(612,64) = 0,69$, $p = ,489$). A l'inverse, il existe un très fort écart

entre la condition *Pas de Communication* et la condition *Communication indirecte* sur cette question ($\Delta M = 1,61$, 95% CI [1,35, 1,87], $t(593,28) = 12,18$, $p < ,001$).

La seule question pour laquelle il existe une différence significative entre les conditions *Communication* et *Communication indirecte* est celle portant sur la justice du traitement de l'employé par son patron. Comme on peut s'y attendre, les participants ont jugé que le patron avait traité son employé de façon moins juste dans la condition *Communication indirecte* que dans la condition *Communication* ($\Delta M = 0,30$, 95% CI [0,11, 0,50], $t(604,62) = 3,05$, $p = ,002$). Il n'y a dans ce cas plus de différence significative entre la condition *Communication indirecte* et la condition *Pas de communication* ($\Delta M = 0,10$, 95% CI [-0,11, 0,32], $t(604,49) = 0,96$, $p = ,339$). Un tel résultat est bien sûr logique, puisque le patron a agi de la même manière dans les deux cas. Les moyennes des différentes conditions sont indiquées dans la figure 15. Ces résultats montrent que les participants ont bien été sensibles à la différence entre les conditions *Communication indirecte* et *Communication*.

Fig. 15. Justice du traitement de la personne punie, expérience 3.



Discussion

L'expérience 3 a montré que la transmission de l'information joue un rôle important dans la perception du concept de punition. A partir du moment où le but de l'infliction d'un dommage est perçu par la personne concernée, alors il semble que l'on peut pleinement parler de punition. Même si l'infliction d'un dommage par une autorité est bien décrite par la plupart des participants comme un cas de punition, cette situation semble moins prototypique qu'une situation où une personne subit un dommage de la part d'une autorité, et comprend que ce dommage a été infligé par cette autorité du fait d'une action négative passée. De plus, l'expérience 3 a confirmé qu'un certain nombre de participants pensent que la communication jouera un rôle éducatif dans le comportement de l'individu puni.

Conclusion

Au début de ce chapitre, nous avons mis en avant cinq questions importantes dans le cadre du lien entre punition et communication. Ces cinq questions sont la question de la légitimité, la question conceptuelle, la question motivationnelle, la question de l'interprétation, et enfin la question du succès éducatif. Comme nous l'avons indiqué au début du chapitre, il existe très peu de données actuellement sur le succès de la punition comme moyen éducatif. Il s'agit évidemment d'une question qu'il sera crucial de résoudre à l'avenir.

En ce qui concerne les quatre autres questions, cependant, les données indiquent assez clairement l'importance de la communication dans la perception de l'acte punitif. Les expériences de Bilz et de Bregant, présentées au début du chapitre, montrent que la peine est interprétée d'une manière symbolique. Les expériences 2 et 3 que nous avons effectuées montrent de leur côté que la communication fait partie du concept prototypique de punition, et que l'infliction d'un dommage possède une plus grande légitimité et est plus satisfaisante si elle passe par une communication.

La question de la motivation est plus compliquée. Comme nous l'avons mis en avant dans notre réplique de l'expérience effectuée par Nadelhoffer, la possibilité de communiquer semble avoir un impact limité sur la motivation à punir tant que la comparaison entre la possibilité de punir sans communiquer et la possibilité de punir en communiquant n'a pas été

rendue saillante au sujet. Cependant, il est possible de mettre en avant le fait que, dans des conditions écologiques, la personne punissant a généralement le choix dans la manière de punir, et a donc concrètement la possibilité de punir ou bien de façon anonyme, ou bien de façon publique. Dans ce cadre, on peut considérer que la communication doit bien jouer un rôle motivationnel dans la vie courante.

Toutes ces données renforcent donc la plausibilité des théories communicationnelles de la peine. Bien sûr, surtout dans le cas de l'examen de la question motivationnelle, il faut souligner les limites des méthodes expérimentales employées, en ce qu'il n'est pas évident que ces méthodes puissent être généralisées aux cas de punition de la vie ordinaire, impliquant notamment des crimes beaucoup plus sérieux que le fait de retirer de l'argent à une tierce personne dans un jeu économique. Ensuite, il faut évidemment mentionner le fait que toutes ces expériences ont été faites auprès de participants américains. Savoir dans quelle mesure ces données seraient également reproduites avec des participants européens est difficile. Cependant, si l'on admet ces limitations, il s'agit bien d'un soutien premier aux théories communicationnelles.

Ce renfort empirique, cependant, ne règle pas toutes les objections contre les théories communicationnelles de la peine, que nous avons examinées au chapitre précédent. On peut toujours objecter que se servir de la peine pour communiquer est barbare, et absurde, et devrait être remplacé par un blâme purement verbal. Nous allons dans le chapitre suivant répondre à ces critiques, et formuler une défense théorique des théories communicationnelles de la peine.

Chapitre 9

Justifier les théories communicationnelles de la peine

Le chapitre précédent nous a permis de montrer que les théories communicationnelles de la peine sont bien ancrées dans les intuitions populaires. Nous allons maintenant procéder à un examen des objections proposées par les opposants aux théories communicationnelles. Comme nous l'avons vu dans le chapitre 7, ces objections sont essentiellement de deux types. La première objection revient à dire qu'il est barbare de faire souffrir quelqu'un pour lui communiquer quelque chose. La deuxième soutient qu'il n'y a pas de nécessité de punir pour communiquer quelque chose à quelqu'un, étant donné qu'une simple information verbale suffirait. Je répondrai ici tour à tour à ces deux objections. Cette réponse nous permettra de mettre en avant le rôle crucial de la récompense et de la peine pour l'institution d'une communauté de valeurs entre individus.

Pourquoi il n'est pas barbare de faire souffrir des individus

Comme nous l'avons vu dans le chapitre 7, une critique courante contre les théories communicationnelles de la peine est de leur reprocher de justifier une pratique barbare, à savoir le fait d'infliger de la souffrance afin de communiquer quelque chose. Cette critique est redoublée par des accusations d'incohérence, mises en avant notamment par Russ Shafer-Landau. Ce dernier reproche en effet aux théories communicationnelles de chercher à jouer un double jeu : elles prétendent agir pour le bien du condamné, tout en lui infligeant de la souffrance, et tout en agissant contre la volonté de la personne punie.

En ce qui concerne le côté barbare de faire souffrir des individus pour leur communiquer quelque chose, cette critique représenterait une objection puissante si jamais la communication était considérée comme étant le seul objectif de la punition, et si la punition devait être faite

pour le bien du condamné. Cependant, comme nous l'avons vu dans le chapitre 7, la communication ne constitue pas la seule justification du fait que nous devons punir les criminels selon leur mérite. La punition sert aussi à satisfaire nos intuitions rétributives, et à prévenir de futurs crimes vis-à-vis de la société. En ce sens, il n'est pas besoin pour une théorie communicative de la peine de justifier l'utilité de la souffrance. Cette justification a déjà été faite. La communication sert uniquement de justification supplémentaire visant à parachever une théorie complète du mérite.

On voit par là également pourquoi les objections de Shafer-Landau ne font mouche que contre un type de théorie éducative très particulière, à savoir, des théories qui justifient la peine uniquement par le fait qu'elle sert le bien de l'individu concerné. Mais il n'est aucunement besoin, dans notre cas, d'aller aussi loin dans l'idée que la peine se fait dans l'intérêt du condamné. Il est tout à fait possible de dire que le dommage infligé au délinquant ne se fait pas avant tout pour son bien, mais pour le bien de la collectivité. Ce bien commun se trouve advenir d'une pluralité de manières, ce qui inclut non seulement des raisons rétributives et dissuasives, mais également des raisons éducatives. L'existence d'une communauté de citoyens éduqués d'un point de vue moral représente avant tout un bien collectif, avant d'être un bien pour chacun des individus concernés. Supposons en effet que le délinquant a reconnu ses erreurs, et accepte l'idée qu'il n'était pas bon de commettre un acte anti-social. Nous pouvons donc parler d'éducation morale. Cette éducation représente-t-elle un bien pour l'ancien délinquant en question ? S'il regrette véritablement ses actions passées, il est probable qu'il est heureux d'avoir changé et d'avoir abandonné ses désirs anti-sociaux. Cependant, il est beaucoup plus naturel de parler ici d'un gain collectif, puisque la communauté a gagné un nouveau citoyen moral en son sein. Même si la théorie communicationnelle s'adresse au délinquant, il n'y a pas de raison de penser qu'elle doit avoir à cœur avant tout le bien de l'individu concerné, plutôt qu'un attachement plus large au bien collectif.

Une telle réponse, qui fait appel au fait que le dommage infligé lors de la punition est justifié de nombreuses manières pourra sembler insatisfaisante au lecteur. Je ferai donc appel à deux autres raisons pour montrer qu'infliger un dommage ou une souffrance au criminel n'est pas un acte barbare. Tout d'abord, il est possible de reprendre un argument défendu par Moore dans sa défense du rétributivisme²³¹. Si nous imaginons une situation où nous commettons par mégarde une grave offense, ou si nous imaginons que nous avons commis un crime sans le faire exprès, alors il semble naturel de penser que nous devrions être puni d'une manière ou d'une

²³¹ M. S. MOORE, « Justifying Retributivism », *op. cit.*

autre, et que cela exprimerait notre sincère repentir. Il est difficile de penser dans ce cas qu'il puisse être barbare de vouloir pour un autre ce que nous voudrions pour nous-même.

Deuxièmement, pour justifier en quoi infliger de la souffrance n'est pas nécessairement barbare, on peut bien entendu faire appel au fait que le délinquant est responsable de ses actes, et qu'il disposait d'une possibilité d'éviter de commettre son délit²³².

Troisièmement, il est évident que juger de savoir dans quelle mesure l'infliction intentionnelle de souffrance est un acte barbare ou non dépend fortement des intuitions individuelles. Dans ce cas précis, les intuitions rétributives possèdent un très fort soutien en général au sein des populations. Un philosophe critiquant des intuitions rétributives ou communicatives peut bien sûr soutenir que les intuitions de la grande majorité de ses concitoyens sont barbares, mais une telle accusation peut sembler perdre fortement de sa force dans un tel contexte.

Réponse aux critiques traditionnelles : pourquoi communiquer

Reprenons la deuxième critique formulée habituellement à l'encontre des théories communicationnelles de la peine. Cette critique consiste à défendre l'idée que le choix de la punition n'est pas approprié pour communiquer quelque chose. Habituellement, la communication passe par des mots ; pourquoi ne pas se contenter d'effectuer une simple remontrance verbale pour critiquer la personne délinquante ?

Face à cette critique, puissante au premier abord, les théories communicationnelles de la peine ont systématiquement répondu de la même manière : la punition est nécessaire pour être sûr que le criminel prenne la punition au sérieux. Sans cela, le délinquant pourrait très bien se contenter de négliger les critiques. Une simple remontrance orale n'est pas assez convaincante : la punition symbolise le degré de gravité de l'infraction. Un bon exemple de ce type de réponse se trouve dans le *On Justice* de Lucas : « L'ignominie est l'expression de la réprobation publique d'un tort. Et pour beaucoup de personnes, dans beaucoup de milieux, il s'agit d'une sanction suffisante. Mais pas pour tout le monde. Certaines personnes sont trop endurcies pour y attacher beaucoup d'importance, et sont tout à fait prêtes à défier l'opinion publique. Sur leur échelle de valeurs ils s'en seront tirés à bon compte, à moins que l'on ne donne une forme

²³² *Ibid.*

tangible à la réprimande dans des termes qui feront sens de leur point de vue. Ils resteront de marbre si le juge les critique simplement. Le but de la punition [...] consiste donc à leur faire comprendre que la réprimande est sérieuse²³³. » Dans ce passage, Lucas exprime l'idée qu'une réprimande simplement orale peut suffire pour certaines personnes, mais pas pour toutes. L'efficacité de la communication pure, sans punition, est donc dépendante du degré auquel une personne attache de l'importance à l'avis d'autres personnes, et à leur bien-être. Il s'agit d'une idée importante sur laquelle nous reviendrons.

La réponse de Lucas semble avoir un certain attrait intuitif, mais il faut bien avouer qu'elle n'est pas parfaitement limpide. Pourquoi donc est-ce que des personnes devraient ne pas prendre au sérieux une réprimande orale ? Est-ce que l'aspect communicatif du dialogue n'est pas développé exactement de la même manière dans le cas d'une communication purement verbale, et dans une punition ? Qu'est-ce que la punition ajoute exactement au sérieux de la communication ?

Pour répondre à ces questions, il faut prendre conscience de l'incertitude dans l'évaluation de la communication, et dans l'importance de mettre en place des procédures parvenant à maximiser la fiabilité des retours que nous nous formulons les uns aux autres. De façon plus générale, je voudrais suggérer que la question de la théorie communicative de la peine doit être interprétée comme faisant partie d'une théorie plus large des pratiques de comptabilité morale. La prochaine section fournira une brève esquisse d'une telle approche.

De l'importance d'une comptabilité morale au sein d'une société

Pour mettre en avant le rôle de la peine dans les pratiques communicationnelles, je voudrais présenter l'esquisse d'une théorie de la comptabilité morale au sein d'une société. Au sein d'un groupe d'individus, afin de maintenir l'équilibre des intérêts entre les différentes personnes, il est important de savoir quelles bonnes et mauvaises actions chaque personne a fait dans le passé. Si l'on veut mettre en place un système d'incitations, alors il est nécessaire de savoir qui a fait

²³³ J. R. LUCAS, *On Justice, op. cit.*, p. 133. Lucas indique également que le lien entre la souffrance et l'indication que l'on prend un tort au sérieux n'est pas limité aux cas de punition, mais concerne également les cas de contribution après un tort involontaire. « Même dans des communautés idéales, il est facile de dire « désolé », mais difficile de le dire du fond du cœur. Afin de prouver sa sincérité, et de montrer sa contrition avec plus que de simples mots, le pénitent ressent donc souvent le besoin de faire un sacrifice substantiel, au-delà de ce qu'il a déjà fait en réparation de ses méfaits. » .

de bonnes choses, afin de pouvoir les récompenser pour cela, et de savoir qui a fait de mauvaises choses, afin de pouvoir leur retirer des bénéfices si nécessaire. L'importance d'un tel système a déjà été établie par nos remarques sur l'importance du mérite pour préserver un système d'incitation efficace (chapitre 6). Il faut donc maintenir une comptabilité exacte des actions effectuées par les différents individus.

Dans ce cadre, en s'inspirant de la théorie de l'échange de responsabilité développée par Chaudhry et Loewenstein, on peut mettre en avant l'idée qu'il existe différents types de communications qui ont pour but de maintenir une comptabilité des actions méritantes effectuées par chaque individu²³⁴. Chaudhry et Loewenstein distinguent quatre types principaux de types de communication, selon que ces derniers concernent des actions positives et négatives, et selon que l'action communicatrice est effectuée par la personne qui a effectué l'action ou la personne qui a reçu l'action. Ces quatre types de communication sont l'auto-congratulation, le remerciement, l'excuse, et le blâme. En ce qui concerne les actions positives, une personne peut se féliciter d'une action qu'elle a faite (auto-congratulation), ou la personne bénéficiaire peut la remercier (remerciement). En ce qui concerne les actions négatives, une personne peut s'excuser de ce qu'elle a fait (excuse), ou une personne victime peut blâmer la personne concernée (blâme).

Dans chacun de ces cas, il est possible de se limiter à une communication passant par la simple parole, ou à un ensemble d'actions qui sont à la fois la traduction et la conséquence de ces actes communicatifs. Dans le cas du blâme, il est possible d'aller plus loin et de punir les personnes concernées ; dans le cas du remerciement, il est possible de récompenser la personne qui a bien agi. De la même manière, dans le cas de l'excuse et de l'auto-congratulation, il est possible pour les personnes ayant causé un dommage ou un bénéfice d'être pro-actives et de traduire ces paroles en acte. Par exemple, une personne qui souhaite s'excuser peut manifester à quel point elle est désolée en présentant un cadeau afin de réparer la relation froissée. Ainsi, chez les Yuqui d'Amazonie étudiés par Allyn Maclean Stearman, une personne se sentant en faute peut chercher à réparer une relation en offrant spontanément du poisson à la personne qui

²³⁴ Chaudhry et Loewenstein formulent cette typologie des manières d'attribuer des actions positives ou négatives à des individus. Cependant, leur théorie se situe dans un tout autre champ de questions, étant donné qu'ils cherchent à analyser les problèmes posés par des situations où un individu a le choix entre blâmer une autre personne ou laisser cette personne s'excuser d'elle-même. Ils analysent alors ce type de situations comme posant à l'acteur un choix entre apparaître froid (s'il blâme l'autre personne) ou risquer d'apparaître incompetent (s'il ne la blâme pas, et que cette deuxième personne ne s'excuse pas). Shereen Jehan CHAUDHRY et George LOEWENSTEIN, « Thanking, Apologizing, Bragging, and Blaming: Responsibility Exchange Theory and the Currency of Communication », *Psychological Review*, à paraître.

a été victime²³⁵. De la même manière, dans nos sociétés, il est possible de témoigner de ses regrets vis-à-vis d'une action néfaste en offrant un cadeau.

Toutes ces pratiques possèdent donc un aspect à la fois purement communicatif, et un aspect communicatif et matériel. L'hypothèse des théories communicatives de la peine, élargie aux pratiques générales de comptabilité morale, est qu'ajouter un aspect matériel joue non seulement un rôle incitatif, mais également un rôle proprement communicatif, et améliore la communication entre les acteurs. Le problème qui se pose à nous est donc celui de savoir en quoi consiste la fonction communicative des transferts matériels.

Pour résoudre ce problème, il est important de distinguer plusieurs fonctions d'un acte communicatif de comptabilité morale. Lorsque je communique un blâme ou un éloge, je communique trois contenus informatifs distincts. Premièrement, je transmets une information relative à un événement objectif qui s'est produit. Par exemple, dans le cas de l'éloge, je rends compte du fait que mon collègue a effectué un travail important. Deuxièmement, je communique le fait que j'ai pris conscience que cet événement s'est produit, et donc que j'agirai à l'avenir en tenant compte de cet événement. Même si ce deuxième aspect est conceptuellement distinct du premier aspect, ces deux fonctions de l'éloge ou du blâme sont pragmatiquement et empiriquement indissociables. Si je transmets une information, j'assume pragmatiquement que j'y crois, et que j'en tiendrai compte à l'avenir²³⁶. Enfin, troisièmement, je transmets une information relative à la manière dont je juge la valeur morale de cet événement. Si j'effectue un éloge enthousiaste, il s'agit d'une indication quant au fait que l'action qui vient d'être effectuée est d'une grande valeur.

L'existence de ces trois contenus informatifs de l'éloge ou du blâme ont une importance cruciale. Il est en effet impossible de coopérer s'il n'existe pas 1) une histoire claire des actes coopératifs effectués par chaque individu, 2) une reconnaissance par chaque individu de ce qu'ont fait d'autres personnes, 3) une communauté de valeur qui permet d'attribuer une échelle d'évaluation des différentes contributions. Les deux premiers points coulent de source. Le troisième point, lié à l'importance de l'existence d'une communauté de valeur et d'appréciation, est plus complexe, et demande une explication.

Un système coopératif repose à la fois sur un accord quant aux faits objectifs, et un accord quant aux évaluations relatives à ces faits objectifs. Si un individu propose un article pour un ouvrage collectif, l'accréditation indique qu'il en est l'auteur. Cela n'est pas en soi suffisant

²³⁵ Allyn Maclean STEARMAN, *Yuqui: Forest Nomads in a Changing World*, Holt, Rinehart and Winston, 1989.

²³⁶ Cette analyse pragmatique de l'assertion remonte au moins aux travaux sémiotiques de Peirce. Charles S. PEIRCE, *Ecrits sur le signe*, Points, 2017.

pour former une communauté d'appréciation. En effet, il existe une incertitude fondamentale quant à la valeur de ce qui a été produit. Il n'y a pas de correspondance directe entre une production matérielle et une appréciation collective. Dans le cas d'un partage des tâches au sein d'un couple, par exemple, les deux membres du couple peuvent attacher une valeur différente à la vaisselle, au ménage, au fait de descendre les poubelles, etc., et peuvent avoir une différence d'appréciation quant à l'importance d'effectuer ces tâches de façon approfondie. Mettre la table peut être considéré ou bien comme une tâche trop triviale pour posséder une véritable valeur, ou peut être évalué comme une véritable contribution à la vie du couple. Un certain nombre de batailles féministes au cours du XXe siècle peuvent ainsi être interprétées comme une lutte pour la réévaluation d'un certain nombre de tâches qui avaient été sous-évaluées dans le passé, comme les tâches domestiques ou la charge d'élever des enfants²³⁷. Dans ce cas, personne ne mettait en doute le fait que ces tâches féminines étaient bien effectuées ; les débats relevaient de l'appréciation subjective de la valeur des contributions féminines. De la même manière, jardiner peut être interprété comme un loisir personnel, ou comme une contribution authentique à l'embellissement de la maison.

Dans ce cadre, il existe toujours une incertitude fondamentale quant à la valeur des contributions des individus. En ce qui concerne des actes négatifs, on retrouve une possibilité similaire d'hésitation. Même dans un cas gravissime, comme le viol, il existe une tendance générale pour les individus inculpés à minimiser la gravité de leur action, notamment en tentant de créer le doute sur le degré de consentement des victimes²³⁸. Dans des cas plus légers, comme le vol, il existe toujours une forte incertitude quant à la valeur des objets ou le degré auquel leurs propriétaires y tenaient.

Cette incertitude dans l'évaluation des actes, positifs ou négatifs, ne serait pas en elle-même un énorme problème si les êtres humains étaient toujours de bonne volonté, et accordaient la même importance à leur prochain qu'à eux-mêmes. Mais comme l'exemple dramatique du blâme des victimes dans le cas du viol le montre bien, ce n'est pas le cas. Le mécanisme de défense de ses propres opinions et intérêts, qu'on appelle raisonnement motivé ou biais en faveur du moi (*myside bias*) en psychologie sociale, indique que nous sommes prêts, consciemment ou inconsciemment, à tordre la valeur réelle des choses afin de préserver nos

²³⁷ Sur les injustices quant à l'évaluation des tâches féminines, voir l'ouvrage classique de Madeleine Guilbert sur la dévalorisation des tâches féminines dans l'industrie. Madeleine GUILBERT, *Les Fonctions des femmes dans l'industrie*, Mouton et Cie, 1966.

²³⁸ Pour un texte classique sur le blâme des victimes, voir Ronnie JANOFF-BULMAN, « Characterological versus behavioral self-blame: Inquiries into depression and rape », *Journal of Personality and Social Psychology*, 1979, vol. 37, n° 10, p. 1798-1809.

intérêts personnels²³⁹. Les recherches en psychologie sociale ont montré de nombreux cas où les intérêts individuels conduisaient à une perception déformée de la réalité. L'exemple classique est celui de l'expérience conduite par les psychologues Hastorf et Cantril auprès d'étudiants des universités de Dartmouth et Princeton en 1951²⁴⁰. Le match de football américain entre les équipes des deux écoles avait donné lieu à une forte controverse. Hastorf et Cantril en ont profité pour demander aux étudiants de chaque université d'estimer le nombre de fautes commises par les membres de chaque équipe. De façon assez prévisible, les étudiants avaient tendance à minimiser les fautes commises par les joueurs de leur équipe, et à exagérer les fautes commises par les joueurs de l'équipe adverse. On voit là une divergence fondamentale sur l'étendue des actes négatifs effectués par deux équipes.

Encore plus important pour notre propos, des travaux récents en psychologie sociale ont insisté sur le fait que nous avons d'autant plus tendance à biaiser notre perception de la réalité dans des cas où il existe une incertitude fondamentale dans l'évaluation de la situation²⁴¹. Pour reprendre l'exemple du match de foot, on peut penser que les supporters des deux équipes seraient d'accord s'il existait une faute particulièrement grave ; dans ce cas, il n'y a pas de possibilité de nier la faute sans tomber dans un grave déni de réalité qui discréditerait le locuteur. Cependant, si la faute est plus ambiguë, alors on voit bien pourquoi il est davantage possible de prétendre qu'il n'y a pas eu faute. La marge d'erreur inévitable dans ce genre de cas conduit au fait qu'un agent ne se discréditerait pas d'un point de vue épistémique s'il nie l'existence d'une faute. Dans ce cadre, plus une évaluation est difficile et ambiguë, plus les agents peuvent faire valoir leurs intérêts personnels tout en maintenant une apparence d'honnêteté.

Or, dans le cas de l'évaluation de la valeur des contributions, on tombe dans un domaine subjectif où l'incertitude la plus totale règne. Il n'existe pas de fait irréfutable quant à la valeur de la contribution d'un employé à une multinationale, quant à l'action d'un chasseur dans une traque collective, ou dans le fait de faire une lessive pour un couple. Dans ces circonstances, la mauvaise foi, ou le raisonnement motivé, peuvent avoir libre cours et briser toute possibilité

²³⁹ Pour un panorama d'exemples et de recherches sur la pensée motivée, on pourra consulter l'article classique de Ziva Kunda. Ziva KUNDA, « The case for motivated reasoning. », *Psychological bulletin*, 1990, vol. 108, n° 3, p. 480-498.

²⁴⁰ Albert H. HASTORF et Hadley CANTRIL, « They saw a game; a case study », *The Journal of Abnormal and Social Psychology*, 1954, vol. 49, n° 1, p. 129-134.

²⁴¹ Jason DANA, George LOEWENSTEIN et Roberto WEBER, « Ethical immunity: How people violate their own moral standards without feeling they are doing so », in David De CREMER et Ann E. TENBRUNSEL (dir.), *Behavioral Business Ethics: Shaping an Emerging Field*, Routledge, 2012, p. 201-219.

d'accord dans l'évaluation de la valeur des choses, chaque partenaire de l'interaction ayant intérêt, ou bien à minimiser, ou bien à maximiser, la contribution de chacun.

Dans ce cadre de doute généralisé sur l'appréciation des contributions, on comprend qu'il soit particulièrement important pour toute société de s'assurer de l'existence d'une communauté de valeurs, et d'une homogénéité dans l'évaluation des actions. Cette idée de l'importance d'une communauté de valeur a été mise en avant de façon particulièrement claire par Adam Smith dans sa *Théorie des sentiments moraux*. « Or, cette harmonie si délicieuse ne peut être obtenue sans une libre communication des sentiments et des opinions. Nous désirons tous, pour cela, sentir de quelle façon nous nous affectons les uns les autres, pénétrer dans les cœurs les uns des autres, et observer les sentiments et les affections qui s'y trouvent réellement. L'homme qui nous permet de ressentir cette passion naturelle, qui nous invite en son cœur, qui pour ainsi dire nous en ouvre toutes grandes les portes, semble pratiquer une espèce d'hospitalité plus délicieuse que toutes les autres. Aucun homme, pourvu qu'il soit d'un bon tempérament ordinaire, ne peut manquer de nous donner du plaisir s'il a le courage d'exprimer ses sentiments réels comme il les sent, et parce qu'il les sent²⁴². » Cette citation montre bien l'incertitude dans les domaines d'évaluation des actes, et l'importance sociale d'un signe tangible de nos appréciations et de nos reproches. La fonction communicative des pratiques de mérite remplit cette fonction.

De l'importance de la reconnaissance et du risque de la méconnaissance

Jusqu'à présent, j'ai présenté une justification des raisons de l'utilité communicationnelle de signes tangibles du mérite. Nous avons besoin de nous repérer dans le labyrinthe des contributions sociales, et de communiquer clairement aux individus des signes d'estime et de mésestime. Bien que cette justification ne soit peut-être pas la justification la plus importante

²⁴² A. SMITH, *Théorie des sentiments moraux*, op. cit., p. 448-449. Il ajoute, p. 177 : « Et nous ne pouvons pas nous satisfaire d'être seulement admirés pour ce qui est admiré chez les autres; nous devons au moins croire que nous sommes admirables pour ce qui les rend admirables. [...] Mais ce bonheur et cette satisfaction sont grandement renforcés quand nous constatons que les autres voient notre conduite et notre caractère précisément comme nous nous étions efforcés de le faire en imagination. Leur approbation confirme nécessairement notre approbation de nous-même. Leur éloge renforce nécessairement le sens que nous avons de mériter l'éloge. Dans ce cas, le désir de mériter l'éloge est si loin de dériver de l'amour de l'éloge que c'est ce dernier qui paraît dériver, du moins dans une grande mesure, du désir d'être digne d'éloge. »

des quatre justifications que j'ai présentées du mérite, il n'en reste pas moins qu'il s'agit d'un trait crucial de notre vie sociale.

Cependant, il serait trompeur de cacher les perversions qui peuvent naître de cette importance communicationnelle de la peine et de la récompense. Étant donné que la récompense est un signe tangible d'estime, et la peine un signe manifeste de réprobation, la peine et la récompense obtiennent une importance qui dépasse de beaucoup leur seule valeur matérielle. Dans ce cadre, toute injustice dans la distribution de la peine et de la récompense obtient une valeur décisive, en tant que l'injustice matérielle est redoublée d'une blessure symbolique.

Axel Honneth est certainement l'auteur contemporain qui a le plus insisté sur la dimension pathologique des injustices dans la distribution des biens matériels. Le passage suivant est particulièrement clair à ce sujet :

« J'ai tout d'abord distingué les analyses portant sur la société dans son ensemble selon qu'elles s'appuient normativement plutôt sur l'idée d'un ordre social juste ou plutôt sur celle d'une réalisation de soi réussie - par « réalisation de soi » il faut entendre un point de repère permettant de s'interroger sur un ordre social pour savoir s'il assure à ses membres des possibilités satisfaisantes de formation de l'identité. A l'aune de cette distinction, j'ai ensuite distingué la philosophie politique de la philosophie sociale ou, pour le dire autrement, les catégories fondamentales de l'injustice des catégories des pathologies sociales. Je pense que la tradition marxiste s'intéressait beaucoup plus aux pathologies sociales qu'aux injustices sociales et l'apport particulier de Marx était peut-être, à ses propres yeux, d'avoir pu démontrer que l'injustice sociale représente simultanément une pathologie sociale. C'est-à-dire qu'il a identifié le fait même de l'injustice, liée à l'exploitation, à une pathologie sociale, qui s'exprime à travers l'aliénation, soit le fait que nous sommes tous, et pas seulement le prolétariat, rendus étrangers à nos propres conditions de vie, à la nature et à nous-mêmes. [...] »

Par « pathologies sociales », j'entends des relations ou des évolutions sociales qui portent atteinte, pour nous tous, aux conditions de réalisation de soi²⁴³. »

Dans ce passage, Honneth met clairement en valeur plusieurs traits qui découlent de l'importance communicationnelle du mérite. Honneth insiste sur le fait que la récompense ou la peine changent les possibilités de réalisation de soi, en tant qu'elles contribuent à former l'identité personnelle. Si je gagne beaucoup d'argent, c'est un signe que je fais une contribution très utile à la société, et donc que j'ai une forte valeur sociale ; si j'en gagne peu, c'est un signe

²⁴³ Axel HONNETH, *La société du mépris : vers une nouvelle théorie critique*, trad. fr. Olivier VOIROL, Paris, La Découverte, 2015, p. 179.

que la société ne valorise pas mon travail. De même, la peine est un signe que ce que j'ai fait est condamné très fortement par la société.

Dans un ordre social juste, ces signes et ces communications contribuent à donner un sens à l'ordre social, en tant qu'il s'agit de signaux véridiques de la valeur des différentes contributions. Dans un ordre social injuste, cependant, il s'agit de sources de perversions, étant donné que les victimes d'injustices subissent la double peine d'un manque matériel et d'une communication faussée.

On peut prendre plusieurs exemples de distributions injustes au sein de nos sociétés, qui ont des conséquences symboliques énormes. Sur le plan pénal, aux États-Unis, la criminalisation des pratiques liées à la drogue a pu conduire à l'incarcération d'une grande partie de la jeunesse noire américaine, associant l'image négative de la condamnation à l'histoire du racisme aux États-Unis. Même si la guerre contre la drogue a été moins dure en France, une bonne partie des peines ces dernières années n'en restent pas moins liées ou au trafic de drogue, ou aux infractions au code de la route.

Sur le plan de la justice distributive, les inégalités titanesques de salaires dans nos sociétés créent également une communication pervertie. L'écart gigantesque entre un salarié payé au SMIC et un cadre dans le domaine de la finance renvoie l'image d'une valorisation sans commune mesure des deux activités. Il n'est en rien évident que ces différences de revenus correspondent à des différences d'utilité sociale. Il existe en conséquence une forme de reconnaissance tronquée, où les individus les mieux reconnus ne sont pas ceux qui ont contribué le plus à la société. La société envoie un retour négatif, incohérent, sur les véritables contributions des individus.

En conséquence, il est possible de tirer deux conclusions normatives de ce risque de méconnaissance. Tout d'abord, puisque la récompense comme la peine jouent un rôle communicatif, il est d'autant plus important que la société réfléchisse aux signaux qu'elle souhaite envoyer. Il est d'autant plus crucial d'organiser la société afin qu'il existe une adéquation entre la communication et le mérite réel des individus.

Deuxièmement, comme l'ont souligné des auteurs comme Michael Walzer ou David Miller, il est nécessaire d'avoir une conception riche du mérite des individus, qui ne se limite pas au pur rôle sur le marché du travail²⁴⁴. Le marché du travail est le domaine où se manifeste le plus clairement la dimension communicative du mérite, du fait de l'aspect tangible de la reconnaissance sociale passant par le salaire. Cependant, il ne s'agit certainement pas de la seule

²⁴⁴ Michael WALZER, *Spheres Of Justice: A Defense Of Pluralism And Equality*, New York, Basic Books, 1983 ; D. MILLER, *Principles of Social Justice*, op. cit.

source de mérite. Chacun peut contribuer à la société aussi bien par un rôle dans le monde associatif, politique, économique, familial. Il est important de garder à l'esprit la pluralité des contributions sociales, et d'accorder une reconnaissance qui ne se limite pas au pur domaine économique. A défaut de la reconnaissance procurée par le salaire, cette reconnaissance élargie pourrait passer par l'attribution de prix ou par une plus grande visibilité d'actions associatives dans la vie publique.

Conclusion

Dans ce chapitre, j'ai offert une réponse aux théories qui considèrent que se servir de la peine ou de la récompense est une pratique complètement superflue pour communiquer la désapprobation ou l'approbation sociale. J'ai notamment mis en avant l'ambiguïté de l'estimation des contributions et des dommages causés par chacun. Afin de préserver la société comme tissu coopératif, j'ai insisté sur l'importance de pratiques publiques, symboliques, de comptabilité morale. Nous devons avoir une claire idée des contributions que nous faisons à la société, et des contributions faites par chaque individu. La récompense comme la peine servent ces fonctions de manifestation publique de l'appréciation des actes de chacun.

Enfin, dans un dernier temps, j'ai insisté sur les risques créés par cette fonction communicative de la récompense et de la peine. Il existe toujours un danger d'attirer l'opprobre sur des actes neutres, si ceux-ci sont punis, et de minorer la contribution de certains individus, si leurs actions ne sont pas récompensées à leur juste valeur. L'impératif de récompenser le mérite en sort renforcé.

Au cours des quatre derniers chapitres, j'ai fourni une justification générale du mérite, notamment en insistant sur la valeur communicative de la récompense et de la peine. Il est maintenant temps de résoudre un des problèmes les plus épineux sur le sujet : le lien entre mérite et responsabilité.

Quatrième partie
Mérite et responsabilité

Chapitre 10

Responsabilité et égalitarisme de la fortune

Les chapitres 6 à 9 ont constitué une défense générale de l'idée de mérite. Une telle justification resterait cependant sans effet si elle n'était pas complétée par le traitement d'un des problèmes centraux de l'étude du mérite : la question de la relation entre mérite et responsabilité.

A première vue, les idées de responsabilité et de mérite semblent étroitement associées. Plus précisément, on pourrait dire que la responsabilité semble une condition indispensable du mérite : pour qu'une personne puisse mériter quelque chose sur la base d'une action, il semble nécessaire que cette personne soit responsable du fait d'avoir effectué cette action. Jusqu'ici, cette idée semble parfaitement naturelle et ne semble pas poser de problèmes particuliers. Bien sûr, on peut douter que nous soyons responsables de quoi que ce soit dans cette vie, si l'on adopte par exemple une position incompatibiliste dure, et que l'on rejette l'existence du libre-arbitre²⁴⁵. Mais, dans ce cadre, l'idée de mérite ne semble pas plus en danger que l'idée générale de liberté ou de responsabilité. La question de l'existence du libre-arbitre est évidemment hautement délicate, et elle ne sera pas traitée ici. Je supposerai qu'une solution générale à l'existence de la responsabilité et du libre-arbitre a été trouvée²⁴⁶.

²⁴⁵ Derk PEREBOOM, *Living without Free Will*, Cambridge University Press, 2001.

²⁴⁶ Pour des exemples de défense d'une existence authentique du libre-arbitre et de la responsabilité, voir par exemple certains auteurs compatibilistes majeurs contemporains comme John Martin FISCHER et Mark RAVIZZA, *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, 1998 ; John Martin FISCHER, *Deep Control: Essays on Free Will and Value*, Oxford, OUP USA, 2013 ; ou des auteurs majeurs incompatibilistes comme Robert KANE, *Free Will and Values*, Albany, State University of New York Press, 1985. J'approuve personnellement la position compatibiliste générale défendue par Fischer, mais je pense que les idées défendues dans ce chapitre peuvent tout aussi bien être associées à une position incompatibiliste libertarienne.

Quelle place pour la responsabilité ? Responsabilité immédiate, régressive, circonstancielle

Pour clarifier le lien entre mérite et responsabilité, il faut impérativement faire la distinction entre plusieurs conceptions de la responsabilité. Plus précisément, je vais mettre en avant la distinction entre trois types de responsabilité, qui correspondent à trois manières différentes d'exercer un contrôle sur son environnement.

Le premier type de responsabilité, le plus simple, correspond à ce que j'appellerai responsabilité immédiate. Ce type de responsabilité traduit le fait que nous avons produit des conséquences dans un contexte où nous disposions d'un contrôle direct sur nos actions et où il était possible de prévoir les conséquences immédiates de ces actions. Autrement dit, si je fais une action A, et qu'il est prévisible que cette action produira des conséquences C, et si le fait de faire l'action A est le produit de mon intention, alors je suis responsable de ces conséquences C²⁴⁷. Une telle responsabilité est évidemment nécessaire pour que l'on puisse dire que je mérite quelque chose. Et, comme le souligne David Miller, ce type de responsabilité est très important pour le mérite : cela permet d'exclure des critères du mérite des actions effectuées par pur hasard, par chance, ou par inadvertance²⁴⁸. Cependant, il ne s'agit pas d'un concept de responsabilité problématique. Sauf à nier le libre-arbitre, il est évident que nous faisons parfois des choses intentionnellement, et que nous maîtrisons parfois les conséquences de nos actes. Dire que le mérite présuppose une responsabilité immédiate est assez évident, et ne pose en rien problème pour la construction d'une théorie du mérite.

A l'opposé de cette conception somme toute très minimale de la responsabilité, il est possible d'adopter une conception beaucoup plus exigeante, qui correspond à ce qui a été appelé la responsabilité régressive²⁴⁹. Même si cette conception n'est pas en elle-même très plausible, il est intéressant de la considérer pour voir à quoi ressemblerait la conception la plus forte possible de l'idée de responsabilité. La responsabilité régressive correspond à l'idée que, pour être responsable de quelque chose, il faut être responsable de l'ensemble des causes de cette chose.

²⁴⁷ Les concepts d'action et d'intention ont évidemment fait l'objet d'une vaste littérature dans la philosophie anglo-saxonne. Là encore, je me contente d'un appareillage conceptuel minimal pour résoudre le problème qui nous occupe, mais le lecteur intéressé pourra se référer à l'anthologie de Marc NEUBERG (dir.), *La Responsabilité : Questions philosophiques*, Paris, Presses Universitaires de France - PUF, 1997 ; ou aux articles classiques de Frankfurt réunis dans Harry G. FRANKFURT, *The Importance of What We Care About: Philosophical Essays*, Cambridge University Press, 1988.

²⁴⁸ D. MILLER, *Principles of Social Justice*, op. cit., p. 143.

²⁴⁹ Susan L. HURLEY, *Justice, Luck and Knowledge*, Harvard University Press, 2005.

On voit bien qu'adopter cette conception conduit à une réaction en chaîne : pour être responsable d'une chose, il faut être responsable des causes de cette chose, ce qui implique d'être responsable des causes de cette cause ; et ainsi de suite. En conséquence, adopter une conception telle que la responsabilité régressive implique que l'on ne peut être responsable de rien si l'on n'est pas responsable de l'univers dans son entier. Même s'il est possible de défendre une telle conception²⁵⁰, je pense qu'il s'agit d'une idée un peu trop baroque pour être vraiment plausible. Adopter une telle conception de la responsabilité revient à déclarer de but en blanc que la responsabilité individuelle est impossible. Sauf à proposer d'autres arguments, je considère qu'il s'agit d'un présupposé inacceptable ; il n'y a pas d'intérêt spécifique à définir un concept de la vie courante de manière à inscrire dans sa définition le fait qu'il s'agit d'un concept impossible à satisfaire.

Passons maintenant au troisième type de conception de la responsabilité, que j'appelle ici responsabilité circonstancielle, et qui peut être interprétée comme une version affaiblie de la responsabilité régressive évoquée au paragraphe précédent. Selon la responsabilité circonstancielle, pour être pleinement responsable d'une chose, il faut avoir joué un rôle causal majeur dans les circonstances qui ont rendu possible l'existence de cette chose ; plus faible est notre rôle causal dans la création des circonstances qui ont rendu notre action possible, plus faible est notre responsabilité. Cette dernière caractérisation est différente de deux manières avec la caractérisation que nous avons faite de la responsabilité régressive. Tout d'abord, l'idée de responsabilité circonstancielle n'exige pas que nous soyons responsables de nos circonstances ; il suffit d'avoir joué un rôle causal important dans leur production. On évite ainsi l'idée de régression à l'infini. Ensuite, l'idée de responsabilité circonstancielle telle qu'elle a été proposée ici admet des degrés. Il ne s'agit pas d'être responsable tout court d'une action ou de ses conséquences ; on est plus ou moins responsable en fonction de la part qu'on a eu dans la production des circonstances qui ont rendu notre action possible. L'idée de responsabilité circonstancielle pose bien davantage de problèmes à une théorie du mérite. En effet, en utilisant un autre vocabulaire, de nombreux philosophes ont défendu l'idée qu'il serait nécessaire de posséder une responsabilité circonstancielle pleine et entière avant de pouvoir dire que l'on mérite quelque chose.

Nous avons donc éliminé une conception peu plausible de l'idée de responsabilité, la responsabilité régressive, et avons mis en avant deux conceptions plus raisonnables, la

²⁵⁰ C'est une stratégie de ce type qu'utilise Galen Strawson dans « The impossibility of Moral responsibility » pour démontrer l'impossibilité de la responsabilité morale. Galen J. STRAWSON, « The Impossibility of Moral Responsibility », *Philosophical Studies*, 1994, vol. 75, n° 1-2, p. 5-24.

responsabilité immédiate et la responsabilité circonstancielle. Armés de cette distinction, nous pouvons maintenant préciser les deux questions principales qui se posent à toute théorie cherchant à lier responsabilité et mérite.

Une question facile, une question difficile : clarifier le problème du lien entre mérite et responsabilité

Il existe deux questions relatives au lien entre mérite et responsabilité, une question facile, et une question difficile. Pour les mettre en avant, prenons le cas d'un maçon qui a effectué une contribution économique dans une entreprise, par exemple, qui a contribué à construire une maison. La question facile est la suivante : faut-il que le maçon possède une responsabilité immédiate vis-à-vis des conséquences de son action pour qu'il puisse être méritant ? Comme nous l'avons vu, la réponse est dans ce cas positive. Une production qui serait un pur coup de chance ne saurait être source de mérite. Entendue dans un sens minimal, la responsabilité est bien une condition nécessaire du mérite. Si l'on admet que la responsabilité et le libre-arbitre existent, et si l'on adopte une conception immédiate de la responsabilité, on sera obligé de considérer que le maçon est responsable de la construction de la maison. Le lien est direct, conscient, évident entre le maçon et le résultat de son action. Il s'agit d'un cas parfaitement transparent de causalité intentionnelle.

Cependant, ce lien clair de responsabilité du maçon vers son résultat n'implique pas que le maçon soit responsable des circonstances qui ont rendu possible son action méritante. Quelles sont les circonstances ayant rendu possible le fait de construire la maison ? D'un point de vue proximal, on peut souligner le fait que le maçon a été embauché pour la construire. D'un point de vue plus distal, il y a le fait que le maçon a reçu une formation professionnelle dans le domaine de la maçonnerie, ce qui lui a permis d'acquérir des compétences absolument nécessaires pour son action. De façon plus distale encore, on peut inclure dans les circonstances qui ont rendu possible cette action méritoire le fait que le maçon a reçu un patrimoine génétique qui l'a peut-être prédisposé à son métier, et le fait qu'il a reçu une certaine éducation et un certain entourage familial. Le même type de causes, plus ou moins proximales ou distales, peuvent se retrouver pour toutes les actions productives possibles ou imaginables. Si l'on imagine un artiste, un homme politique, un scientifique, un banquier, un enseignant, un artisan,

un ouvrier, il se retrouvera forcément des causes plus ou moins lointaines qui auront rendu possible cette action productive.

La question difficile portant sur la relation entre mérite et responsabilité est alors la suivante : est-il nécessaire que quelqu'un ait eu un rôle causal majeur dans la production des circonstances qui l'on rendu méritant pour que l'on puisse dire que le mérite ait un sens ? Cette question est particulièrement délicate parce qu'il est évident que nous n'avons pas joué de rôle causal important dans la production d'une grande partie des choses qui ont pu nous rendre méritants. Nous sommes fondamentalement projetés dans ce monde avec de certaines dispositions et un certain environnement familial, et ces deux types de circonstances, génétiques et environnementales, ont une influence décisive sur nos actions ultérieures. Si une conception saine du mérite exigeait que nous ayons eu un rôle majeur dans la causation de nos talents et de notre milieu familial, alors le mérite serait impossible.

Reprise des conditions d'Olsaretti

Pour affiner notre étude sur le lien entre mérite et responsabilité, il faut revenir aux conditions d'une théorie raisonnable du mérite proposées par Serena Olsaretti, que nous avons étudiées dans le chapitre 1. Dans ce chapitre, nous avons défendu l'idée que la plupart des conditions proposées par Olsaretti étaient indispensables pour la formulation d'une théorie intéressante du mérite. Ces conditions incluaient par exemple l'idée que le mérite devait posséder une valeur sans être réductible à un autre principe de justice, ou que le mérite devait satisfaire les critères de justice comparative. Nous avons laissé de côté une condition substantielle défendue par Olsaretti, parce que nous considérons que cette condition n'était en rien évidente. La condition posée par Olsaretti est la suivante : une conception plausible du mérite présuppose qu'il existe une juste égalité des chances d'acquérir du mérite. Olsaretti défend ainsi l'idée que, si l'on veut dire qu'un individu mérite de gagner davantage qu'un autre, il faut au minimum que les deux personnes concernées aient eu à l'origine les mêmes chances de devenir productif et d'accomplir des actes méritants²⁵¹.

Selon la conception défendue par Olsaretti, il faudrait donc que deux individus aient en face d'eux les mêmes possibilités a priori, et que toute différence de mérite entre ces deux individus

²⁵¹ S. OLSARETTI, *Liberty, desert and the market*, op. cit., p. 24.

s'expliquent par des choix que ces deux individus ont fait. Dans ce cadre, si les deux individus ont par exemple devant eux les mêmes possibilités d'étude, mais que l'un d'entre eux choisit de faire des études longues et l'autre des études courtes, et que la personne qui a choisi de faire des études longues obtient en fin de compte un meilleur salaire que l'autre, alors on pourra dire que cet individu méritait d'obtenir un meilleur salaire. Si, à l'inverse, l'individu ayant fait de plus longues études disposait d'avantages dont ne jouissait pas l'autre personne (par exemple du fait d'un milieu social plus favorable, ou parce qu'il possédait de plus grandes facilités naturelles), alors il n'est pas possible de dire que cette personne mérite de gagner plus que l'autre.

En un mot, Olsaretti considère que le mérite ne fait aucun sens sans l'existence d'une juste égalité des chances. Une telle condition est beaucoup plus plausible que l'idée que nous devrions être responsables de l'existence de l'univers tout entier pour pouvoir mériter quoi que ce soit. En effet, les conditions posées par Olsaretti n'excluent pas *a priori* l'existence du mérite. C'est donc cette condition dont nous ferons l'examen au cours de ce chapitre. Est-il nécessaire que deux individus aient disposé des mêmes opportunités pour que l'on puisse dire que l'un est plus méritant que l'autre ?

De façon absolue, la condition posée par Olsaretti semble fausse. Nous avons pris dans le premier chapitre l'exemple de Miller d'un individu qui en sauve un autre de la noyade. Nous ne disposons pas en général d'une égalité des chances dans le fait de nous promener près d'autres individus en train de se noyer. Il s'agit d'un événement rare. Pourtant, il semble naturel de dire que le sauveteur mérite les remerciements, et que d'autres personnes non-impliquées dans le sauvetage n'en méritent aucun, alors même qu'elles auraient probablement agi de la même façon si elles en avaient eu la possibilité.

Une telle objection n'est cependant pas décisive. La manière la plus simple de sauver la proposition d'Olsaretti est de dire que les individus en question ne doivent pas disposer de possibilités de faire exactement les mêmes choses, mais simplement de possibilités de faire des choses également méritoires. Si un individu n'a pas eu la possibilité de sauver quelqu'un de la noyade, il a peut-être eu la possibilité de faire du volontariat dans une association, de s'occuper de personnes en difficulté, etc. Dans ce cas, les deux individus disposent de possibilités semblables, même s'ils ne disposent pas des mêmes possibilités exactement. La condition de juste égalité des chances peut alors être analysée comme une condition globale régissant l'ensemble des opportunités dont disposent les individus au cours de leur vie, et ne devrait pas s'appliquer à chaque occasion particulière. Savoir si une telle condition reste plausible sous cette forme demande un plus long examen.

Trois questions cruciales

Sur la base de l'analyse proposée par Olsaretti, nous pouvons mettre en avant trois questions cruciales qu'il s'agit de résoudre ici. Première question : une juste égalité des chances d'être méritant est-elle une condition nécessaire pour l'attribution du mérite ? Si cette condition possède à première vue une assez forte plausibilité, il faut remarquer qu'elle pose une très grande difficulté à toute théorie du mérite. Il semble impossible d'atteindre jamais une parfaite égalité des chances d'être méritant ; il semble qu'il y aura toujours, *a minima*, des inégalités persistantes qui feront qu'une certaine partie de la population disposera de circonstances plus favorables qu'une autre. Rien que l'existence de différentes familles, fournissant à chaque enfant une éducation plus ou moins différente, rend très douteuse l'existence d'une égalité stricte des opportunités offertes à chaque individu. Selon l'idée courante, au moins depuis la parution de la *Théorie de la justice*, une parfaite égalité des chances semble impossible à moins d'abolir la famille. La condition stricte posée par Olsaretti revient à dire que le mérite ne fait jamais vraiment sens, sauf à imaginer une société dystopique où la famille serait éliminée et où chaque enfant vivrait exactement dans le même environnement que tous les autres.

Dans ce contexte, il faut se poser une deuxième question, qui découle directement de la première. Quel est l'impact de circonstances injustes et d'inégalités dans la capacité des individus à être méritants ? En d'autres termes, est-ce que des circonstances inégales annulent complètement le mérite, le diminuent, ou n'ont aucun impact sur lui ? Si l'on imagine des différences de circonstances très faibles entre deux individus, il semble douteux que ces différences faibles puissent avoir véritablement pour effet d'annuler complètement le mérite. Par exemple, imaginons deux personnes issues de milieux sociaux très semblables, mais où l'une des deux personnes est issue d'un milieu légèrement plus favorisé. Si l'on laisse tous les autres facteurs égaux par ailleurs, il semble évident qu'un revenu parental très légèrement supérieur (mettons de 5% plus grand dans un cas que dans un autre) constitue un avantage, mais un avantage qui semble trop ténu pour justifier l'élimination complète du mérite. L'importance d'une juste égalité des chances pour la détermination du mérite ne semble pas être une question de tout ou rien, mais une question de graduation.

Une troisième question possède une importance à peu près aussi grande : comment comprendre l'idée d'égalité des chances d'être méritant ? Parmi l'ensemble des facteurs qui influencent la capacité des individus à être méritants, lesquels doivent être considérés comme des facteurs légitimes, lesquels comme des facteurs illégitimes ? Dans le cas d'un milieu social

plus favorable, il semble assez évident que ce facteur doit être considéré comme illégitime, en tant qu'un plus grand mérite ne devrait pas, idéalement, découler de quelque chose qui ne semble pas propre à l'individu. Un cas beaucoup plus difficile semble être la question des talents naturels²⁵². De nombreux théoriciens modernes ont objecté que les talents naturels ne devraient pas être considérés comme une source légitime de mérite, en tant que les individus n'en sont en rien responsables. Une telle idée a été développée avec force dans les théories contemporaines de l'égalitarisme de la fortune, notamment à partir des travaux pionniers de Ronald Dworkin au début des années 1980²⁵³. Cependant, la nature injuste d'une différence de talents semble beaucoup moins claire que l'injustice causée par des différences de milieux sociaux.

La complexité de la légitimation d'inégalités par le biais des talents naturels prend un aspect d'autant plus important que de nombreux théoriciens du mérite se sont opposés aux égalitaristes de la fortune en défendant la validité des talents naturels comme facteurs d'inégalités légitimes. De fait, l'idée de méritocratie, telle qu'elle a été introduite (de façon critique) par Young dans son livre de 1958 *The Rise of Meritocracy*, décrit la méritocratie avant tout comme une hiérarchie basée sur l'intelligence, comprise comme étant déterminée essentiellement sur des facteurs génétiques²⁵⁴. Bien sûr, le livre de Young est une satire. Cependant, des défenseurs contemporains de l'idée de mérite, comme David Miller ou David Schmidtz, ont considéré que les talents naturels constituent un facteur parfaitement légitime pour mettre en avant le fait que certains individus sont plus méritants.

Ceci ne veut bien sûr pas dire que les défenseurs du mérite ont fait une apologie sans partage du mérite fondé sur des talents naturels. Les auteurs ont régulièrement mis en avant l'importance de la responsabilité individuelle dans la justification des inégalités²⁵⁵. Mais dans ce cas, si la responsabilité est si importante pour justifier la présence d'inégalités, il reste à

²⁵² Par talents naturels, je ne veux bien sûr pas dire que des individus seraient équipés dès la naissance de capacités parfaites en aucun domaine. L'idée de talents naturels renvoie ici simplement à une plus grande probabilité, d'origine génétique, de développer certaines capacités ou de faire certains types d'apprentissages. Comme cela a été évoqué dans l'appendice du chapitre 4, il est bien établi d'un point de vue scientifique qu'il existe une influence génétique sur les comportements humains, et les capacités humaines. Cela ne veut bien sûr en rien dire que les influences sociales ne sont pas importantes. La nature comme l'environnement jouent un rôle majeur. Eric TURKHEIMER, « Three Laws of Behavior Genetics and What They Mean », *Current Directions in Psychological Science*, 2000, vol. 9, n° 5, p. 160-164 ; Robert PLOMIN, John C. DEFRIES, Valerie S. KNOPIK et Jenae M. NEIDERHISER, « Top 10 Replicated Findings from Behavioral Genetics », *Perspectives on psychological science : a journal of the Association for Psychological Science*, 2016, vol. 11, n° 1, p. 3-23 ; Valerie S. KNOPIK, Jenae M. NEIDERHISER, John C. DEFRIES et Robert PLOMIN, *Behavioral Genetics*, 7^e édition., New York, Worth Publishers, 2016.

²⁵³ Ronald DWORKIN, *Sovereign Virtue: The Theory and Practice of Equality*, Harvard University Press, 2002.

²⁵⁴ M. YOUNG, *The Rise of the Meritocracy 1870-2033*, op. cit.

²⁵⁵ Voir par exemple D. MILLER, *Principles of Social Justice*, op. cit.

expliquer pourquoi le fait qu'il n'existe pas une égalité des chances d'acquérir des talents naturels ne pose pas plus de problèmes aux tenants du mérite.

La place de la responsabilité : le débat autour de l'égalitarisme de la fortune

Ce débat autour d'une juste égalité des chances et sur l'importance de la responsabilité a possédé une très grande importance dans les débats contemporains en philosophie politique, notamment depuis 1981 et la publication par Ronald Dworkin d'une série d'articles sur la nature de l'égalité. Les idées de Dworkin ont été rapidement reprises et défendues par une série d'auteurs égalitaristes, notamment par les philosophes Richard Arneson, G. A. Cohen et John Roemer²⁵⁶. Même si ces auteurs ont chacun défendu des théories différentes, ils ont été regroupés sous le titre d'égalitarisme de la fortune depuis la publication en 1999 de l'article critique fondateur d'Elizabeth Anderson, « What is the Point of Equality ?²⁵⁷ ».

Les différents égalitaristes de la fortune se retrouvent dans l'approbation de l'importance de la responsabilité individuelle pour comprendre la valeur de l'idée d'égalité, et dans l'insistance sur l'importance de la neutralisation du hasard pour la justice sociale. Le cœur de l'égalitarisme de la fortune tient dans l'idée qu'il faut égaliser les opportunités ouvertes aux individus, et non pas le résultat final que ceux-ci obtiennent. Dans ce cadre, même si deux individus disposent des mêmes chances au départ, il est toujours possible qu'ils se retrouvent avec des biens différents à l'arrivée. Ces résultats inégaux peuvent être la conséquence de deux processus considérés comme légitimes par les égalitaristes de la fortune. Tout d'abord, si ces différences à l'arrivée s'expliquent par des choix que ces individus ont faits, alors il est légitime que les biens des individus reflètent leurs propres choix. Ensuite, il est également possible que ces différences s'expliquent par le fait que ces individus ont pris des risques, ou fait des paris

²⁵⁶ Richard J. ARNESON, « Equality and Equal Opportunity for Welfare », *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 1989, vol. 56, n° 1, p. 77-93 ; G. A. COHEN, *On the Currency of Egalitarian Justice, and Other Essays in Political Philosophy*, Princeton, Princeton University Press, 2011 ; John E. ROEMER, *Equality of Opportunity*, Harvard University Press, 2000.

²⁵⁷ Elizabeth S. ANDERSON, « What Is the Point of Equality? », *Ethics*, 1999, vol. 109, n° 2, p. 287-337. Dworkin a depuis refusé l'appellation. Il existe de nombreuses différences entre les théories proposées par ces auteurs ; une des différences les plus importantes renvoie à la question de la nature de ce qui doit être égalisé. Dworkin propose par exemple une égalisation des ressources des individus, tandis qu'Arneson propose une égalisation du bien-être des individus. Étant donné que cette question est accessoire pour notre propos, et étant donné que ces différences entre les différents égalitaristes de la fortune n'ont pas de conséquence pour ce qui va suivre, je traiterai à l'avenir les différents égalitaristes de la fortune comme s'ils constituaient un bloc cohérent.

librement consentis. Par exemple, imaginons que ces deux individus ont parié à la loterie, et supposons que l'un d'entre eux ait gagné le gros lot, tandis que l'autre a tout perdu. Étant donné que ces individus étaient entrés librement dans ces paris, même si la différence entre les deux individus s'explique avant tout par la chance, cette chance n'est pas injuste parce qu'elle est la pure conséquence de choix librement consentis. Une même analyse peut être conduite vis-à-vis d'un certain nombre d'autres choix risqués dans la vie quotidienne, comme le fait de conduire sans ceinture ou de fumer.

Les égalitaristes de la fortune acceptent donc qu'un certain nombre d'inégalités apparentes peuvent ne pas être des inégalités réelles, étant donné qu'elles peuvent découler de choix ou de prises de risques conscients. Cela ne veut bien sûr pas dire que les égalitaristes de la fortune considèrent que toute redistribution est exclue, ou que les inégalités présentes sont globalement justifiées. En l'état actuel des choses, les égalitaristes de la fortune considèrent qu'une grande partie des inégalités ne découlent pas des choix ou d'un hasard consenti (« option luck »), mais découlent d'un hasard brut (« brute luck ») qui s'impose aux individus. Les exemples les plus importants pris par les égalitaristes de la fortune d'inégalités découlant d'un hasard brut proviennent de la classe sociale de naissance, des talents naturels, ou des maladies ayant des facteurs génétiques. Un individu ne choisit pas le milieu social dans lequel il naît, pas plus qu'il ne choisit son patrimoine génétique. Dans ce cadre, toute inégalité découlant de ces facteurs ne provient pas des choix individuels, constitue une rupture de l'égalité, et doit être compensée.

On voit directement l'attrait qu'ont pu avoir les théories des égalitaristes de la fortune. En effet, cette théorie représente une combinaison de deux idées associées habituellement à des bords politiques différents, la responsabilité et l'égalité. De plus, il s'agit d'une théorie qui préconise une forte redistribution sociale, sans pour autant considérer que cette redistribution doit être illimitée. Cependant, à partir des années 1990, une série d'auteurs égalitaristes se sont élevés pour rejeter cette théorie, en considérant que l'égalitarisme de la fortune constitue une violation des valeurs égalitaristes démocratiques²⁵⁸. Je vais présenter ici brièvement les arguments de la principale critique, Elizabeth Anderson.

Anderson rejette les arguments des égalitaristes de la fortune parce que ceux-ci se fondent sur l'idée que l'égalité doit être une égalité distributive. En d'autres termes, les égalitaristes de la fortune considèrent qu'il existe un certain type de biens (les chances d'obtenir du bien-être, par exemple) qui doit être distribué également. Anderson rejette cette focalisation sur un type

²⁵⁸ *Ibid.* ; Samuel SCHEFFLER, « What Is Egalitarianism? », *Philosophy & Public Affairs*, 2003, vol. 31, n° 1, p. 5-39 ; Jonathan WOLFF, « Fairness, Respect, and the Egalitarian Ethos », *Philosophy & Public Affairs*, 1998, vol. 27, n° 2, p. 97-122.

de bien. Selon elle, le but de l'égalitarisme est d'assurer un certain type de relation entre individus, à savoir une relation caractérisée par une égalité de statut. Cette focalisation sur l'égalité de statut conduit Anderson à rejeter l'attention que les égalitaristes de la fortune accordent à la responsabilité. En effet, selon Anderson, le plus important vis-à-vis d'une inégalité distributive n'est pas de se demander si les individus sont responsables ou non de son apparition. Le plus important est de savoir si cette inégalité menace le statut des citoyens qui en sont victimes. Par exemple, si certains citoyens sont tombés dans une grande pauvreté ou n'ont plus de quoi satisfaire leurs besoins fondamentaux, il est, selon Anderson, accessoire de se demander dans quelle mesure ils sont responsables de leur sort. L'essentiel est de se rendre compte que leur statut met en danger l'égalité de statut et de respect entre citoyens, et de se rendre compte ainsi qu'il faut leur venir en aide²⁵⁹.

Au risque de paraître tiède, je dois concéder que les deux camps possèdent des arguments forts en leur faveur, et je préférerais adopter une position pluraliste en la matière. Comme le défend Anderson, parvenir à maintenir une égalité de statut entre les différents citoyens est un objectif politique fondamental, et l'on peut considérer que le maintien de cette égalité est plus important que le traçage exact des responsabilités de chacun. Même si une personne dans le dénuement extrême est partiellement responsable de son sort, cela ne constitue pas un argument fort pour ne pas l'aider. Une telle perspective, qui accepte la primauté de la satisfaction des besoins et du respect prodigué à chaque citoyen, a été défendue plus récemment par certains égalitaristes de la fortune, comme Kok-Chor Tan²⁶⁰. Ce dernier défend en conséquence l'importance de la responsabilité uniquement une fois qu'un certain seuil de satisfaction des besoins a été atteint.

Je suivrai sur ce plan la perspective générale de Kok-Chor Tan, en tant qu'il semble effectivement nécessaire de concilier responsabilité, satisfaction des besoins, et égalité de respect entre citoyens. Ma critique de l'égalitarisme de la fortune passera par d'autres voies que celles défendues par Anderson. Contre les égalitaristes de fortune, je défendrai l'idée que, même s'il semble important de prendre en compte la responsabilité des individus, la responsabilité

²⁵⁹ Anderson fait plusieurs autres critiques à l'encontre des égalitaristes de la fortune. Une critique importante est par exemple le fait que les égalitaristes de la fortune ont tendance à mettre sur le même plan les inégalités naturelles et les inégalités sociales, alors qu'Anderson défend une conception institutionnelle qui nie qu'il puisse exister des inégalités naturelles. D'autres critiques ne sont pas exactement fondamentales, et visent essentiellement à montrer que l'égalitarisme de la fortune peut difficilement être appliqué. Par exemple, Anderson considère que les mesures visant à évaluer le degré de responsabilité de chaque individu, et leurs talents naturels, constitueraient une mesure intrusive et dégradante. Il ne s'agit pas d'une objection fondamentale, en tant que les égalitaristes de la fortune peuvent répondre que la meilleure manière de mettre en place leur théorie ne passerait pas par des actions dégradantes.

²⁶⁰ Kok-Chor TAN, *Justice, Institutions, and Luck: The Site, Ground, and Scope of Equality*, Oxford, Oxford University Press, 2014.

individuelle ne constitue pas un facteur si important que d'autres facteurs ne doivent jouer un rôle dans les questions de justice sociale. En particulier, je défendrai l'idée que nous devons accepter que des individus puissent être plus méritants parce qu'ils ont effectué une contribution plus importante, et ce, même s'ils ne sont pas ultimement responsables des raisons qui les ont rendus plus méritants que d'autres. Un individu qui est prêt à acquérir plus facilement des capacités, du fait de dispositions innées, peut mériter des récompenses davantage qu'un autre, alors même que seules des dispositions génétiques différentes les séparaient au départ.

Cependant, cette justification de l'importance des dispositions innées dans une analyse de la justice sociale ne saurait conduire à justifier entièrement toutes les inégalités découlant de dispositions naturelles différentes. En effet, nous disposons de solides raisons de tenir à l'égalité. Comme les égalitaristes de la fortune, je pense que les facteurs tenant à la responsabilité individuelle constituent les raisons les plus fortes pour se départir d'un égalitarisme strict. Si les inégalités découlant de différences de talents naturels peuvent justifier des inégalités supplémentaires, il n'y a pas de raison de penser que de très fortes inégalités puissent être justifiées de cette manière. Un argument de ce type sera davantage développé dans le chapitre 12.

Notre position consiste donc à défendre qu'une juste égalité des chances n'est pas une condition nécessaire pour que l'on puisse parler de mérite, ou pour dire qu'un individu mérite de gagner davantage qu'un autre. Dans le cas particulier des inégalités de talent, je défends donc l'idée qu'un individu peut être davantage méritant sur la base de faits dont il n'est ultimement pas responsable.

Une telle proposition pourra sembler discutable au lecteur. En effet, l'ensemble de ce travail s'est fondé sur l'idée qu'il est important de se baser sur les intuitions ordinaires pour défendre une théorie plausible du mérite. Or, on pourrait considérer que les intuitions courantes partent précisément de l'idée qu'une juste égalité des chances est une condition fondamentale du mérite. L'importance de la responsabilité individuelle est en effet fondamentale dans les intuitions ordinaires relatives au mérite. De nombreux philosophes, à commencer par Olsaretti, ont souligné le caractère indiscutable de l'idée que le mérite doit se fonder sur la responsabilité²⁶¹. Dans le prochain chapitre, je présenterai différentes études expérimentales qui montreront clairement que le sens commun ne considère pas que le mérite doit nécessairement se fonder sur une juste égalité des chances. Les intuitions courantes sont donc compatibles avec l'existence du mérite, même dans des situations caractérisées par des inégalités de départ dans

²⁶¹ Cf. les auteurs cités par Freiman et Nichols dans Christopher FREIMAN et Shaun NICHOLS, « Is Desert in the Details? », *Philosophy and Phenomenological Research*, 2011, vol. 82, n° 1, p. 121-133.

la capacité à être méritant. Dans le chapitre 12, je fournirai une justification générale de ces idées courantes, et du fait que le mérite garde son sens non seulement dans des situations d'inégale capacité à contribuer, mais même dans des circonstances où cette inégale capacité à contribuer est causée par des facteurs injustes. Le mérite est une idée trop importante pour que des circonstances globalement injustes lui fassent perdre toute valeur.

Chapitre 11

Responsabilité et mérite : études expérimentales

Dans le chapitre 10, nous avons vu qu'il existe des débats intenses autour de la relation entre mérite et responsabilité. Plus précisément, ces débats ont porté autour de deux questions. Tout d'abord, autour de la question de savoir dans quelle mesure une juste égalité des chances d'être méritant est nécessaire pour que le mérite prenne sens. Ensuite, autour de la question de savoir si des facteurs arbitraires d'un point de vue moral (comme l'existence d'inégalités de talent naturel) peuvent jouer un rôle causal majeur dans la création d'inégalités légitimes. Ce chapitre constituera un examen empirique de la question. Après un examen de la littérature en psychologie expérimentale sur le sujet, je présenterai deux expériences personnelles où j'examine la question plus en détails. Comme nous le verrons, ces expériences montrent qu'une grande partie des gens considère que des inégalités légitimes peuvent naître de facteurs arbitraires, comme des différences de talents naturels, alors même qu'il existe un soutien très fort à l'idée d'égalité des chances. Une justification normative de ces intuitions courantes aura lieu dans le chapitre 12.

La perception de la génétique et de l'essence

Avant de s'interroger sur la perception de la légitimité des talents naturels et des capacités socialement acquises, il faut d'abord se demander dans quelle mesure les individus pensent que les gènes ont un impact sur le comportement ou les capacités humaines. Si les individus ne pensaient pas que les gènes peuvent avoir une influence sur la capacité à acquérir des capacités socialement valorisées, s'interroger sur la possibilité d'un effet légitimant des facteurs

génétiques resterait une question assez abstraite²⁶². Heureusement, les vingt dernières années ont vu le développement d'un champ d'étude nouveau particulièrement pertinent pour notre propos : l'étude des perceptions populaires des influences génétiques.

D'après un certain nombre d'études, les personnes tendent à considérer que les êtres vivants possèdent une essence particulière, qui implique qu'ils possèdent un ensemble de traits stables tout au long de leur vie. Un aspect important de ce schéma essentialiste tient à la détermination génétique de certains traits particuliers. Comme plusieurs études l'ont montré, les personnes tendent à attribuer des différences génétiques entre des groupes sociaux à partir du genre ou de l'ethnicité de ce groupe²⁶³. Selon une revue de la littérature effectuée par Dar-Nimrod et Heine en 2011²⁶⁴, la perception de différences génétiques entre des groupes sociaux conduit les personnes à inférer que ces différences sont immuables, que les gènes ont un pouvoir explicatif majeur pour rendre compte des différences entre groupes, et que les groupes sociaux avec un profil génétique spécifique sont homogènes et différents des autres groupes.

Cependant, malgré l'existence d'un nombre considérable d'études portant sur les perceptions de différences de *groupes* relatives à des traits génétiques, la recherche portant sur la perception de différences génétiques entre *individus* reste relativement limitée. Il existe néanmoins un article récent de bonne qualité sur le sujet. Dans deux études récentes portant sur des échantillons larges, Willoughby et ses collègues ont montré que les individus possèdent généralement une croyance forte dans la transmission génétique de traits physiques et psychologiques (comme la couleur des yeux, l'obésité, la dépression, ou la personnalité), et que les estimations d'héritabilité effectuées par le public étaient en général relativement en accord avec la compréhension scientifique de l'héritabilité des traits²⁶⁵. L'intelligence, et deux autres talents (l'athlétisme et les capacités musicales) ont été, en moyenne, perçus comme étant partiellement héréditaires, même s'il y avait une très large variation entre individus en termes de jugements de l'importance de facteurs génétiques.

Il y a donc bien un intérêt à interroger les gens sur la perception de la légitimité d'inégalités découlant de talents naturels. L'idée qu'il existe de plus grandes capacités tirées de facteurs

²⁶² Les quelques paragraphes suivants sont partiellement repris d'un article en préparation, écrit avec Geoff Goodwin.

²⁶³ Nick HASLAM, Louis ROTHSCHILD et Donald ERNST, « Essentialist beliefs about social categories », *British Journal of Social Psychology*, 2000, vol. 39, n° 1, p. 113-127 ; Nick HASLAM, Brock BASTIAN, Paul BAIN et Yoshihisa KASHIMA, « Psychological Essentialism, Implicit Theories, and Intergroup Relations », *Group Processes & Intergroup Relations*, 2006, vol. 9, n° 1, p. 63-76. Bien sûr, le fait que les gens attribuent un fondement génétique à des différences de groupe n'implique pas que ces perceptions soient correctes.

²⁶⁴ Ilan DAR-NIMROD et Steven J. HEINE, « Genetic Essentialism: On the Deceptive Determinism of DNA », *Psychological Bulletin*, 2011, vol. 137, n° 5, p. 800-818.

²⁶⁵ Emily A. WILLOUGHBY, Alan C. LOVE, Matt MCGUE, Jack QUIGLEY et James J. LEE, « Free will, determinism, and intuitive judgments about the heritability of behavior », *Open Science Framework*, 2018.

génétiques correspond bien à des intuitions courantes sur le sujet. Cela ne veut bien sûr pas dire que les individus considèrent que des différences de talents constituent une source légitime d'inégalités. Sur ce plan-là, les rares résultats expérimentaux obtenus ont été plus partagés.

De l'importance des talents naturels, et des capacités socialement acquises

Depuis 2011, deux études ont été publiées qui ont commencé à examiner la perception de la légitimité d'inégalités découlant de différences de talents naturels. En 2011, Christopher Freiman et Shaun Nichols publient un article intitulé « Is Desert in the Details? »²⁶⁶. Leur article est entièrement fondé sur l'idée qu'il existerait un écart important entre les représentations concrètes et abstraites que nous nous faisons à propos du mérite. Selon les auteurs, dans les cas concrets, nous admettons parfaitement l'importance des talents pour déterminer ce que mérite un individu. Par exemple, si nous voyons que quelqu'un a fait une contribution importante, nous considérons qu'il doit obtenir une récompense, et ne nous préoccupons pas tellement de savoir dans quelle mesure cette contribution est due à des efforts ou à des talents importants. À l'inverse, dans l'abstrait, nous aurions une forte méfiance vis-à-vis de l'idée que des choses dont nous ne sommes pas responsables puissent avoir un impact sur notre capacité à mériter quelque chose.

Pour tester cette hypothèse, Freiman et Nichols se sont servis de vignettes décrivant ou bien une condition concrète (deux jongleurs ou deux chanteuses possédant des différences de talent naturel, conduisant à des revenus inégaux) ou bien une condition abstraite (dans laquelle les gens indiquaient dans quelle mesure il pouvait être juste que des personnes gagnent davantage du fait d'avantages génétiques). Dans l'expérience de Freiman et Nichols, les personnes ont considéré que l'inégalité de revenus était juste dans le cas de la condition concrète, mais injuste dans le cas de la condition abstraite. La différence entre les deux conditions était considérable (une différence de 3 points sur une échelle allant de 1 à 7, 7 indiquant un revenu plus juste).

Plus récemment, Goya-Tocchetto et ses collègues ont tenté de répliquer l'expérience de Freiman et Nichols, mais avec plusieurs innovations, qui ont considérablement changé les

²⁶⁶ C. FREIMAN et S. NICHOLS, « Is Desert in the Details? », *op. cit.*

résultats²⁶⁷. Tout d'abord, Goya-Tocchetto a non seulement considéré l'impact des différences de talent naturel, mais aussi l'impact de capacités socialement acquises, et dérivées de différences de milieu social. Ce dernier type d'étude nous intéresse tout particulièrement afin de déterminer dans quelle mesure les personnes jugent que le mérite peut garder un sens et une validité, même dans des conditions sociales injustes. Ensuite, les auteurs ont cherché à éliminer des facteurs confondants à l'œuvre dans l'expérience originale de Freiman et Nichols. Dans la condition abstraite utilisée par Freiman et Nichols, le texte était le suivant : « Supposez que certaines personnes gagnent plus d'argent que d'autres uniquement parce qu'ils possèdent des avantages génétiques. » Ce texte n'est absolument pas clair. Il n'est en effet pas évident de savoir dans quelle mesure ces avantages génétiques facilitent un travail de qualité, qui est le facteur proximal à partir duquel l'on peut dire qu'un individu mérite de gagner un salaire. On peut ainsi penser que des personnes ont considéré que cet « avantage génétique » pouvait être un facteur parfaitement injuste ou arbitraire, comme la couleur de peau ou la beauté des individus²⁶⁸. En conséquence, Goya-Tocchetto et ses collègues ont proposé de nouvelles vignettes abstraites qui précisent clairement que les individus favorisés génétiquement gagnent davantage du fait d'un travail de plus grande qualité.

Les vignettes abstraites utilisées par Goya-Tocchetto et ses collègues prenaient en conséquence la forme suivante : « Imaginez que des scientifiques travailleurs gagnent plus d'argent que d'autres scientifiques tout aussi travailleurs, uniquement à cause du fait qu'ils possèdent des avantages génétiques qui les rendent plus talentueux d'un point de vue intellectuel. » A l'inverse, les vignettes concrètes prennent l'exemple de deux scientifiques précis, et donnent beaucoup plus de détails sur leurs études et leur motivation.

A partir de ces vignettes modifiées et améliorées, les auteurs ont obtenu les résultats suivants. Tout d'abord, ils ont répliqué l'effet principal trouvé par Freiman et Nichols, à savoir un moins grand accord avec l'influence des talents naturels sur le mérite dans une formulation abstraite que dans une formulation concrète, mais avec une taille d'effet bien moindre. Dans leur expérience, les participants étaient plutôt d'accord avec l'idée que la personne ayant des talents naturels plus importants méritait de gagner davantage que l'autre, aussi bien dans la condition abstraite que dans la condition concrète (avec une moyenne d'accord au-dessus de 4,5 et 5, pour les conditions abstraite et concrète respectivement).

²⁶⁷ Daniela GOYA-TOCCHETTO, Matthew ECHOLS et Jen WRIGHT, « The lottery of life and moral desert: An empirical investigation », *Philosophical Psychology*, 2016, vol. 29, n° 8, p. 1112-1127.

²⁶⁸ De fait, j'ai effectué une réplique de l'expérience de Freiman et Nichols, que je ne présenterai pas ici, où plusieurs personnes ont justifié leur réponse indiquant qu'ils interprétaient le texte comme référant spécifiquement à une discrimination raciale basée sur la couleur de peau.

Par ailleurs, les auteurs ont trouvé que les inégalités découlant de milieux sociaux différents étaient considérées comme étant beaucoup moins légitimes que les inégalités découlant de différences de talents naturels. Dans toutes les conditions, que la formulation soit abstraite ou concrète, les individus ont jugé que les inégalités découlant de milieux sociaux étaient illégitimes.

Expérience 1

J'ai décidé de procéder à une réplique extension de l'expérience de Goya-Tocchetto et de ses collègues pour plusieurs raisons. Tout d'abord, bien sûr, je voulais m'assurer de la robustesse des résultats expérimentaux. Ensuite, il me semblait crucial de pouvoir comparer directement les intuitions des gens relatives à une situation d'égal salaire et à une situation de salaire inégal. En effet, les expériences proposées par Freiman, Nichols, Goya-Tocchetto et leurs collègues ont toujours présenté aux participants une situation d'inégalité, et leur ont demandé dans quelle mesure ils trouvaient juste une telle situation. Le problème avec cette méthode est qu'elle ne permet pas de comparer directement la justice de différents états de choses possibles : une absence d'inégalité, une faible inégalité, une forte inégalité. J'ai essayé de remédier à ce défaut en sollicitant des jugements relatifs à différents types possibles d'égalité ou d'inégalité.

Enfin, j'ai cherché à corriger ce que je percevais comme un biais dans les vignettes utilisées par Goya-Tocchetto et ses collègues. En effet, pour étudier les perceptions quant à la justice d'inégalités découlant de milieux sociaux différents, Goya-Tocchetto et ses collègues ont utilisé la vignette suivante :

« Imaginez qu'Amy et Beth veulent toutes les deux être architectes. Elles font toutes les deux autant d'efforts dans leurs études. Les parents de Beth sont plus riches et ont la possibilité de payer pour qu'elle aille à une meilleure école et à une meilleure université que celles que les parents d'Amy sont capables d'offrir à leur fille. Uniquement en conséquence de ce désavantage social, le parcours scolaire de Beth est beaucoup plus impressionnant. En conséquence, Beth obtient un meilleur travail et gagne plus d'argent qu'Amy²⁶⁹. »

²⁶⁹ D. GOYA-TOCCHETTO, M. ECHOLS et J. WRIGHT, « The lottery of life and moral desert », *op. cit.*

Le problème avec une vignette de ce type est que les différences de milieu social décrites ne débouchent sur aucune différence pertinente sur le marché du travail. Un parcours scolaire impressionnant ne semble en rien indiquer que Beth est une meilleure architecte qu'Amy. En d'autres termes, une interprétation raisonnable de cette vignette est qu'elle demande s'il est juste que, à travail de qualité égale, Beth soit payée davantage qu'Amy. Or, il semble assez évident que les origines sociales découlent sur des inégalités injustes si elles ne conduisent qu'à des privilèges. La question intéressante, d'un point de vue du mérite et d'un point de vue psychologique, est de savoir dans quelle mesure des inégalités dues à une contribution plus importante sont considérées comme injustes dans le cas où cette capacité de contribution plus importante est due à des origines sociales différenciées. C'est cette question que j'étudierai dans la présente expérience.

Participants

J'ai recruté 100 participants sur *Amazon Mechanical Turk*. Après avoir éliminé 5 participants qui ont fourni des justifications indiquant qu'ils ne prenaient pas l'expérience au sérieux (par exemple, « A3KF4IP2MUS8QQ »), j'ai obtenu un échantillon de 95 participants. 45 se sont identifiés comme des hommes, et 50 comme des femmes. La moyenne d'âge était de 36,38 ans (Ecart-Type: 11,22).

Méthode

Je me suis servi d'un plan expérimental croisé (*within-subject design*), où les participants ont eu accès à toutes les conditions expérimentales. Toutes les vignettes indiquaient que plusieurs individus différaient en capacités productives pour des raisons qui n'étaient pas liées à des choix et à des efforts, et demandaient aux participants de juger dans quelle mesure il était juste que ces personnes obtiennent un même salaire, un salaire légèrement différent, ou un salaire extrêmement différent.

Les conditions variaient en fonction de l'origine des inégalités de capacités productives (origines sociales ou génétiques) et en fonction du caractère abstrait ou concret des vignettes.

Dans la condition abstraite et génétique, les participants pouvaient lire la vignette suivante : « Imaginez que certains scientifiques travailleurs font un travail de meilleure qualité que d'autres scientifiques également travailleurs, et ce, uniquement parce qu'ils disposaient d'avantages génétiques qui les ont rendus plus talentueux d'un point de vue intellectuel²⁷⁰. » Dans le cas de la condition abstraite et sociale, la fin de la vignette devenait « ... et ce, uniquement parce qu'ils disposaient d'avantages sociaux comme une famille aimante et une meilleure éducation qui leur ont donné davantage de chances de développer leurs capacités intellectuelles. »

Dans le cas des vignettes concrètes, la vignette génétique reprenait essentiellement celle développée par Goya-Tocchetto et ses collègues. Le texte était le suivant : « Imaginez qu'Amy et Beth veulent toutes les deux devenir des ingénieurs informatiques. Elles font toutes les deux autant d'efforts dans leurs études. Même si les maths représentent le talent naturel le plus important d'Amy comme de Beth, Beth arrive toujours à trouver des solutions plus efficaces pour ses programmes informatiques qu'Amy, à cause de différences dans leur constitution génétique. Uniquement à cause de cet avantage génétique, la programmation de Beth est beaucoup plus impressionnante. »

Dans le cas de la condition concrète et sociale, j'ai modifié le texte utilisé par Goya-Tocchetto et ses collègues afin d'éliminer les facteurs confondants. Afin de varier les vignettes et de limiter l'ennui des participants, le métier en question était l'architecture et non l'informatique. Alors que la vignette est globalement identique à celle que le lecteur a lu plus haut, elle se terminait de la façon suivante : « Uniquement à la suite de cet avantage social, Beth a la possibilité d'apprendre beaucoup plus de choses et de devenir une meilleure architecte qu'Amy. » Avec cette fin modifiée, il devient clair que Beth fait un travail de meilleure qualité qu'Amy, et que son avantage social s'est transformé en de plus grandes capacités personnelles.

Pour forcer les participants à comparer les situations d'égalité ou d'inégalités de salaire, je leur ai demandé d'indiquer leur accord (sur une échelle de 1 à 7) avec les trois phrases suivantes :

« Ce serait juste si ces scientifiques avanta­gés d'un point de vue génétique gagnaient exactement le même revenu que les autres. »

²⁷⁰ La profession indiquée renvoyait au fait d'être scientifique afin de pouvoir préciser que les avantages génétiques étaient pertinents pour le travail en question. Goya-Tocchetto et ses collègues se sont servis de plusieurs types de professions différentes ; toutes les professions utilisées ont donné des résultats similaires, ce qui explique que je me sois concentré sur le cas de scientifiques.

« Ce serait juste si ces scientifiques privilégiés d'un point de vue génétique gagnaient un peu plus d'argent que les autres. »

« Ce serait juste si ces scientifiques privilégiés d'un point de vue génétique gagnaient beaucoup plus d'argent que les autres. »

Même si le texte exact de ces phrases variait en fonction des conditions, il gardait toujours cette même structure. Enfin, j'ai demandé aux participants de justifier leurs réponses, d'indiquer leur âge, leur genre, leur niveau d'étude, et leurs attitudes politiques.

Résultats

Tout d'abord, les résultats obtenus répliquent la différence entre conditions concrètes et abstraites que Freiman et Nichols avaient obtenue, puis que Goya-Tocchetto et ses collègues avaient répliquée. Dans le cas du jugement portant sur des différences de salaires faibles comme fortes, les participants considéraient que de telles différences de salaire étaient plus justes dans la condition concrète que dans la condition abstraite. Les différences entre les deux conditions, abstraite et concrète, tournaient toujours autour de 0,40 points (sur une échelle allant de 1 à 7, 7 représentant un plus fort accord). Ces différences étaient hautement significatives dans tous les cas²⁷¹. Comme dans le cas de la réplication de Goya-Tocchetto et de ses collègues, la différence est bien moindre que celle originellement trouvée par Freiman et Nichols.

²⁷¹ Dans le cas de la condition naturelle, les participants étaient davantage d'accord avec l'idée que les personnes privilégiées méritaient de gagner un peu plus d'argent dans la condition concrète que dans la condition abstraite : $M_d = 0,43$, 95% CI [0,17, 0,69], $t(94) = 3,32$, $p = ,001$. Toujours dans le cas de la condition naturelle, ils étaient également davantage d'accord avec l'idée que ces personnes privilégiées devraient gagner beaucoup plus d'argent dans la condition concrète que dans la condition abstraite : $M_d = 0,44$, 95% CI [0,15, 0,74], $t(94) = 2,97$, $p = ,004$.

Ces résultats se retrouvent presque à l'identique dans le cas des conditions où les différences ont une origine sociale. Dans ce cas également, les participants étaient davantage d'accord avec l'idée que les personnes privilégiées méritaient de gagner un peu plus d'argent dans la condition concrète que dans la condition abstraite : $M_d = 0,34$, 95% CI [0,08, 0,59], $t(94) = 2,61$, $p = ,011$. Même si ce dernier résultat n'est pas significatif, il est dans la même direction que les autres résultats, et est en ce sens probablement un faux négatif. Dans le cas de la question portant sur une forte différence de salaire, la différence est significative, et indique toujours un plus fort accord dans le cas de la condition concrète : $M_d = 0,41$, 95% CI [0,13, 0,69], $t(94) = 2,90$, $p = ,005$.

Cependant, je n'ai pas trouvé de différence significative entre conditions concrètes et abstraites dans le cas de la question portant sur le même revenu, même si les résultats tendaient vers une plus forte aversion à l'inégalité dans le cas des conditions abstraites que concrètes. Dans le cas de la condition génétique, l'accord est légèrement plus faible pour la condition concrète que pour la condition abstraite : $M_d = -0,15$, 95% CI [-0,44, 0,14], $t(94) = -1,01$, $p = ,315$. On obtient un résultat similaire dans le cas de la condition sociale : $M_d = -0,09$, 95% CI [-0,41, 0,22], $t(94) = -0,59$, $p = ,555$.

J'ai également répliqué les différences trouvées par Goya-Tocchetto et ses collègues et ce qui concerne l'écart entre conditions sociales et génétiques. Aussi bien dans les conditions abstraites que concrètes, l'accord avec la justice d'inégalités de revenus était plus fort dans le cas des origines génétiques que dans le cas des origines sociales. L'écart entre les deux conditions était toutefois, là aussi, plus faible que ce qui avait été trouvé par Goya-Tocchetto et ses collègues, résultat probable de l'élimination des facteurs confondants dans le cas des vignettes décrivant les différences sociales. Ainsi, dans le cas des conditions abstraites comme concrètes, les différences tournaient autour de 0,40 points (sur une échelle de 1 à 7, toujours), indiquant une différence non négligeable, mais pas aussi forte que les différences de 1 à 1,5 trouvées dans les expériences de Goya-Tocchetto²⁷².

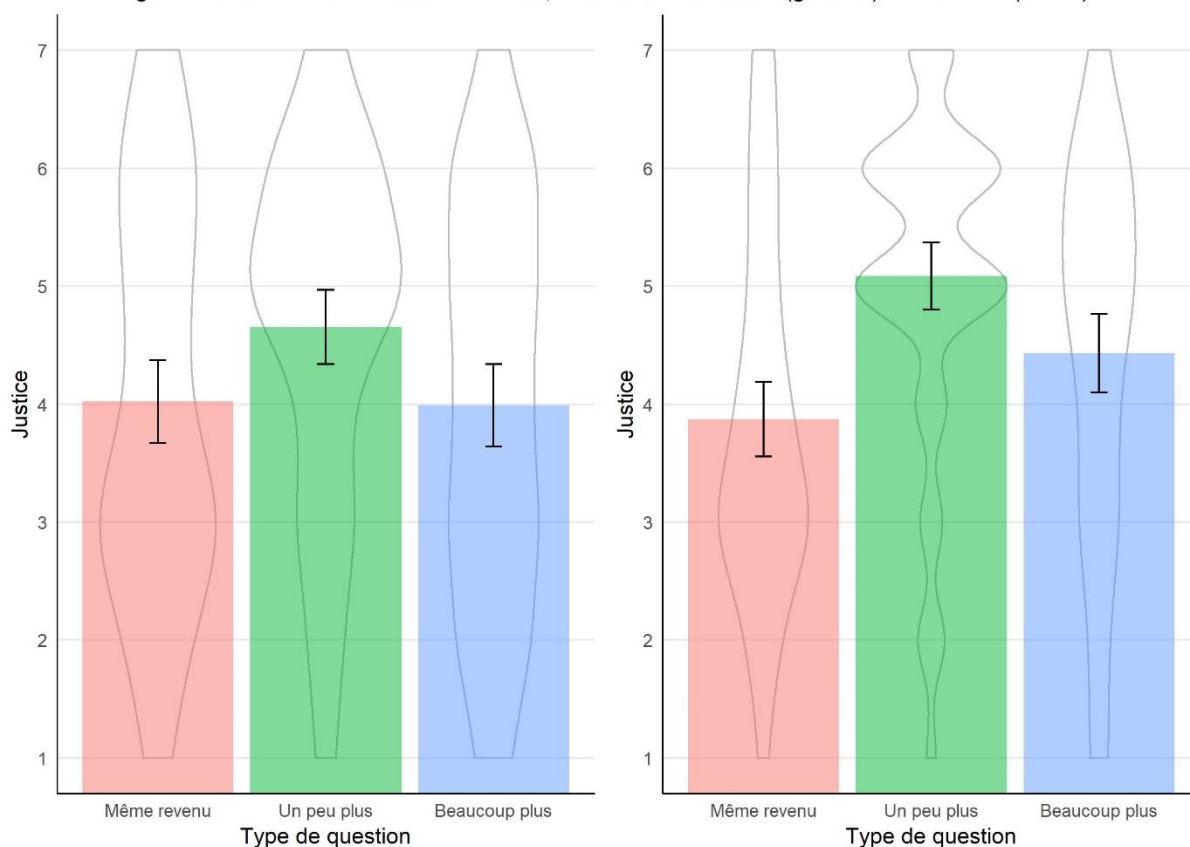
Passons maintenant à l'examen des différences entre les jugements de justice relatifs à un salaire égal, comparé à un salaire inégal. Commençons par les conditions où les différences de productivité sont dues à des facteurs génétiques. Dans ce cas, aussi bien dans les conditions abstraites que concrètes, les participants ont jugé que la situation la plus juste serait une situation d'inégalités limitées. L'accord moyen avec la phrase « Ce serait juste si ces scientifiques avantagés d'un point de vue génétique gagnaient un peu plus d'argent que les autres. » était en effet significativement plus important que l'accord avec l'idée qu'il serait juste s'ils gagnent le même salaire, aussi bien dans le cas de la condition abstraite ($M_d = -0,63$, 95% CI [-1,19, -0,07], $t(94) = -2,24$, $p = ,027$) que dans le cas de la condition concrète ($M_d = -1,21$, 95% CI [-1,72, -0,70], $t(94) = -4,73$, $p < ,001$). La différence est assez importante dans le cas de la condition abstraite (0,63), et devient énorme dans le cas de la condition concrète (1,21). Le lecteur peut voir ces différences dans la figure 16. Une faible différence de salaire était aussi considérée comme plus juste qu'une forte différence de salaire dans le cas des conditions abstraites ($M_d = -0,66$, 95% CI [-0,93, -0,40], $t(94) = -4,94$, $p < ,001$) et concrètes ($M_d = -0,65$, 95% CI [-0,87, -0,44], $t(94) = -6,06$, $p < ,001$).

²⁷² Les résultats sont très similaires pour toutes les questions, et pour toutes les conditions : une plus grande tolérance pour l'inégalité dans le cas des différences d'origine naturelle. Ainsi, dans le cas de la condition abstraite, une égalité de salaire était considérée comme moins juste dans la condition naturelle que dans la condition sociale : $M_d = -0,35$, 95% CI [-0,64, -0,06], $t(94) = -2,37$, $p = ,020$. Un même résultat était obtenu dans le cas de la condition concrète : $M_d = -0,40$, 95% CI [-0,67, -0,13], $t(94) = -2,98$, $p = ,004$.

De la même manière, dans la condition abstraite, un salaire légèrement supérieur était considéré comme plus juste dans la condition naturelle que dans la condition sociale : $M_d = 0,36$, 95% CI [0,08, 0,64], $t(94) = 2,54$, $p = ,013$. Un même résultat se retrouvait dans le cas de la condition concrète : $M_d = 0,45$, 95% CI [0,24, 0,67], $t(94) = 4,20$, $p < ,001$.

Enfin, une forte différence de revenu était aussi considérée comme plus juste dans le cas de différences naturelles dans le cas de la condition abstraite : $M_d = 0,41$, 95% CI [0,14, 0,68], $t(94) = 3,00$, $p = ,003$. Le résultat était similaire dans le cas de la condition concrète : $M_d = 0,44$, 95% CI [0,19, 0,70], $t(94) = 3,42$, $p = ,001$.

Fig. 16. Justice dans la condition Nature, formulation abstraite (gauche) et concrète (droite).

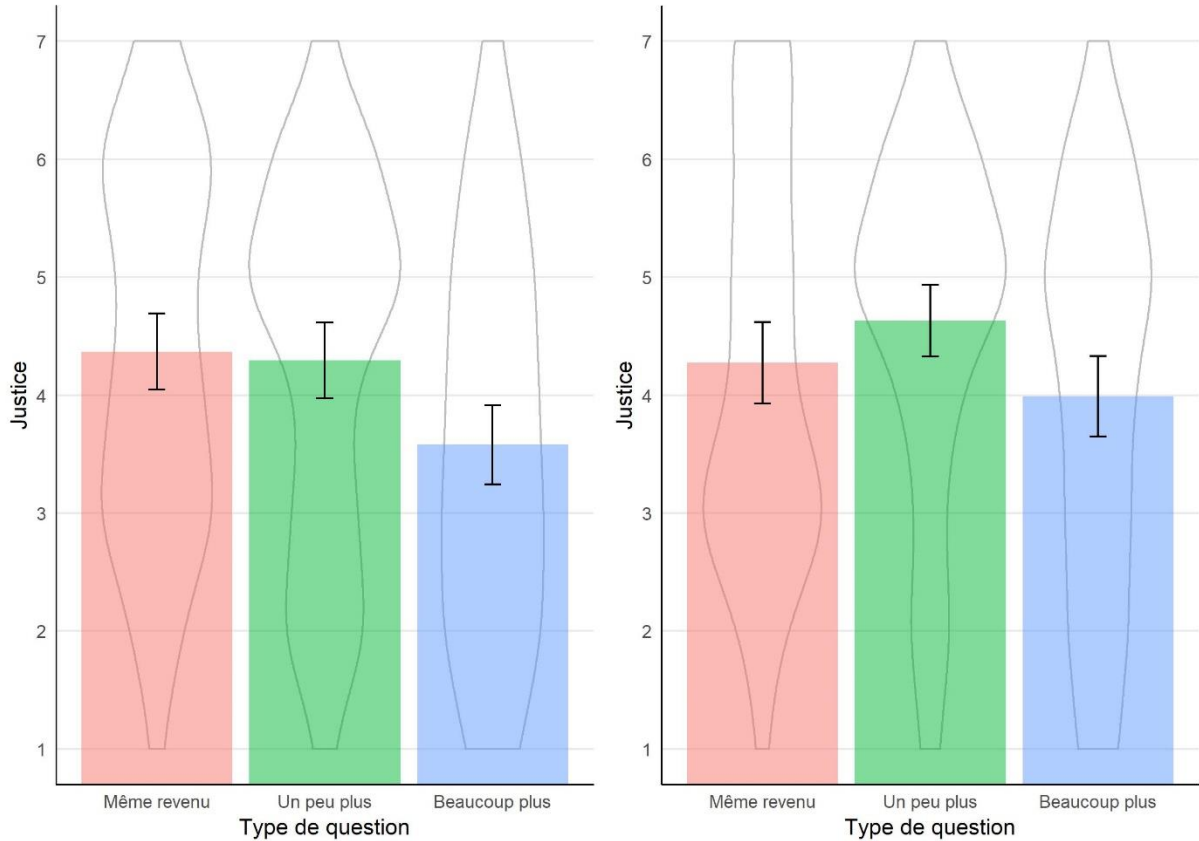


Nous venons donc de voir que les participants ont globalement trouvé que, dans le cas des inégalités d'origine naturelle, la condition la plus juste était la condition d'inégalités limitées. La situation est assez différente en ce qui concerne les inégalités d'origine sociale. Dans ce cadre-là, les inégalités limitées sont aussi très fortement préférées aux inégalités fortes, aussi bien dans la condition abstraite ($M_d = -0,72$, 95% CI $[-0,94, -0,50]$, $t(94) = -6,47$, $p < ,001$) que dans la condition concrète ($M_d = -0,64$, 95% CI $[-0,86, -0,42]$, $t(94) = -5,79$, $p < ,001$). Cependant, il n'existe plus de préférence claire pour des inégalités limitées plutôt qu'une situation d'égalité salariale. De plus, et c'est là un résultat particulièrement intéressant, il n'existe pas non plus de forte préférence pour une situation d'égalité salariale par rapport à une situation de légères inégalités²⁷³. Avec des moyennes tournant autour de 4,4 dans la condition abstraite et de 4,7 dans la condition concrète, il semble n'y avoir aucun consensus sur

²⁷³ Les différences ne sont pas significatives, et, dans la condition abstraite, l'intervalle de confiance des différences entre les deux conditions n'inclut pas de différence majeure. Cela ne veut bien sûr pas dire qu'il n'existe pas de différence entre les deux conditions. La différence dans la condition abstraite est la suivante : $M_d = 0,07$, 95% CI $[-0,44, 0,59]$, $t(94) = 0,28$, $p = ,778$. La différence dans la condition concrète est : $M_d = -0,36$, 95% CI $[-0,92, 0,21]$, $t(94) = -1,26$, $p = ,211$.

le fait qu'une légère différence de salaire soit injuste, alors même que la différence de productivité avait des origines que les participants ont probablement dû juger comme étant parfaitement injustes. Dans le cas concret, la moyenne des réponses est même significativement au-dessus du point neutre établi à 4 ; les participants ont semblé juger une légère différence de salaire comme étant plutôt juste, même si son origine était injuste.

Fig. 17. Justice dans la condition Société, formulation abstraite (gauche) et concrète (droite).



Discussion

Cette expérience permet de tirer plusieurs conclusions sur la perception des talents naturels ou des avantages sociaux, qui toutes vont dans la direction du fait qu'une responsabilité totale n'est pas nécessaire pour que les personnes jugent qu'une inégalité de salaire puisse être justifiée.

Dans le cas d'inégalités de productivité ayant leur origine dans des différences génétiques, les personnes jugent qu'un salaire légèrement supérieur est plus juste qu'un salaire parfaitement égal. Ce choix se retrouve aussi bien dans le cas d'une formulation abstraite que concrète, même

si la différence est beaucoup plus forte dans le cas de vignettes concrètes. Il semble donc que l'absence de responsabilité vis-à-vis des talents naturels ne change pas le fait que ces talents puissent conduire à des inégalités de salaire jugées méritées.

Plus surprenant, les participants ont été divisés quant aux jugements relatifs à la justice d'inégalités salariales découlant de milieux sociaux différents. Dans le cas des vignettes concrètes, les jugements ont même montré que les participants jugeaient une légère inégalité comme étant plutôt juste. Il semble que l'attachement au principe qu'un meilleur travail justifie un meilleur salaire est tel qu'il parvient à compenser le fait que la capacité à effectuer un meilleur travail découlait ultimement de facteurs injustes.

Cet attachement au mérite découlant d'inégalités de productivité est cependant à relativiser. En effet, dans toutes les conditions, les participants ont préféré une situation de légères inégalités à une situation de fortes inégalités. En l'absence de choix ou de responsabilité individuelle forte, il semble que les différences de talents ne peuvent que contribuer minimalement à la perception de la justification des inégalités.

Cette première expérience ne nous apprend cependant que très peu par rapport aux préférences des individus quant à ce que serait une société idéalement juste. Nous avons vu que, si un individu a acquis, pour des raisons injustes dues à des inégalités sociales, de plus grandes capacités qu'un autre individu, alors une grande partie de la population jugera qu'il est plutôt juste que cette personne gagne davantage que la personne qui n'a pas été ainsi favorisée. Mais il est probable que la plupart des gens préféreraient qu'il existe une égalité des chances, et qu'ils favoriseraient une société où les individus ne disposeraient pas de chances aussi inégales à la suite de leurs origines sociales différentes.

En d'autres termes, ces expériences indiquent qu'une grande partie des gens favorisent l'idée que les talents naturels et des capacités socialement acquises peuvent justifier de faibles inégalités. Une fois que ces inégalités de productivité sont en place, un certain nombre de personnes pensent qu'elles justifient des inégalités. Cependant, cela ne veut pas dire que les gens ne favoriseraient pas idéalement une société dans laquelle il n'existerait pas d'inégalité de productivité. On pourrait imaginer en effet que la société perçue comme idéale serait une société où règnerait une totale égalité des chances, c'est-à-dire où tout le monde disposerait au départ des mêmes probabilités d'accéder à n'importe quelle position sociale, quels que soient les talents naturels ou le milieu social d'origine. Pour étudier ces attitudes par rapport à l'élimination des inégalités de productivité non-choisies, j'ai conduit une seconde étude, visant à déterminer les opinions relatives à la compensation des inégalités de milieux sociaux ou de talents naturels.

Expérience 2

Pour examiner les perceptions des différences de talents naturels et de capacités socialement acquises, j'ai construit des vignettes présentant une société imaginaire où le gouvernement aurait décidé de lancer un plan éducatif visant à remédier aux inégalités naturelles ou sociales. Plus précisément, dans une condition, dite *Différences sociales*, le plan éducatif vise à éliminer les inégalités produites par les différences de milieu familial, qui conduisent certains enfants issus d'un milieu favorisé à développer leurs talents de façon bien plus importante que ce qui a lieu pour des enfants issus de milieux moins favorisés. Dans une seconde condition, dite *Différences naturelles*, le plan éducatif cherche à éliminer les différences entre enfants naturellement plus doués et enfants naturellement moins doués. Dans les deux cas, l'élimination des différences de capacités se fait par un soutien massif aux enfants les moins favorisés, et la politique du gouvernement est un succès.

A priori, il serait possible de faire deux prédictions opposées. Tout d'abord, il serait possible que les participants jugent que les inégalités de capacités d'origine naturelle et sociale sont fondamentalement aussi injustes l'une que l'autre, et dans ce cas favoriseraient la mise en place d'un plan égalitariste aussi bien pour les cas d'inégalités naturelles que sociales. Mais il serait également possible que les participants ne jugent pas que les inégalités de productivité dues à des différences de talents naturels soient injustes, et qu'ils jugent le plan de façon beaucoup moins favorable que dans le cas des inégalités d'origine sociale²⁷⁴. Dans cette expérience, je teste ces deux prédictions opposées. Comme nous allons le voir, les participants ont jugé en général le plan éducatif de façon beaucoup plus positive dans le cas des inégalités sociales que dans le cas des inégalités naturelles.

²⁷⁴ Un rejet du plan dans le cas des inégalités d'origine sociale serait très surprenant, étant donné ce que nous savons sur l'importance des intuitions relatives à l'égalité des chances. Cf. Shai DAVIDAI et Thomas GILOVICH, « Building a More Mobile America—One Income Quintile at a Time », *Perspectives on Psychological Science*, 2015, vol. 10, n° 1, p. 60-71.

Participants

Nous avons recruté 99 participants sur *Amazon Mechanical Turk*. Après avoir éliminé 6 participants qui ont fourni des justifications indiquant qu'ils ne prenaient pas l'expérience au sérieux (par exemple, « Nope »), nous avons obtenu un échantillon de 93 participants. 45 se sont identifiés comme des hommes, et 48 comme des femmes. La moyenne d'âge était de 33,10 (Ecart-Type : 11,23).

Méthodes

Selon un plan expérimental croisé (*within-subject design*), les participants ont reçu tour à tour deux conditions, une condition *Différences naturelles* et une condition *Différences sociales*. Le texte des deux scénarios était le suivant :

« Dans le pays de Komland, il existe un système important d'éducation publique, qui encourage une bonne éducation pour les enfants dès un très jeune âge.

[Différences naturelles] En conséquence, il n'y a pas d'inégalité des chances basée sur des facteurs sociaux dans cette société. Les enfants de toutes les classes sociales ont les mêmes chances de réussir. Cependant, cela ne veut pas dire que tous les enfants réussissent de la même manière, étant donné que certains sont naturellement plus doués que d'autres.

[Différences sociales] Cependant, cela ne veut pas dire que les enfants de toutes les classes sociales ont les mêmes chances de réussir. Étant donné que les enfants issus de familles aisées tendent à obtenir un environnement positif, des cours privés, et un meilleur accès à des activités culturelles, ils tendent à développer leurs talents de façon bien plus étendue que les enfants issus de milieux désavantagés. »

Après avoir lu chacun de ces scénarios, les participants apprenaient l'existence d'un plan éducatif gouvernemental : « Après un long débat, pour corriger ces inégalités, le gouvernement de Komland a décidé de mettre en place un nouveau plan éducatif. En conséquence de ce plan, les enfants obtiennent des cours privés, payés par l'État. Les enfants issus de milieux désavantagés reçoivent davantage de ces cours privés que les enfants issus de milieux favorisés [*Différences naturelles* : Moins un enfant est doué, plus il reçoit de ces cours particuliers.] Le coût de cette politique est important, mais celle-ci compense complètement les différences de

capacités entre enfants de différents milieux sociaux [les différences de capacités naturelles entre enfants]. En conséquence, quand ils ont atteint l'âge de 20 ans, tous les membres de Komland finissent par avoir des talents égaux ou comparables, quelle que soit leur classe sociale d'origine [quels que soient leurs talents à l'origine]. »

A la fin de chaque condition, les participants devaient indiquer leur accord, sur une échelle de 1 à 7, avec les phrases suivantes :

« Mettre en place le plan et donner des cours particuliers aux enfants de milieux défavorisés était une décision juste.

Komland aurait été une société juste, même si elle n'avait pas décidé de mettre en place le plan.

Il est juste de dépenser plus d'argent pour des enfants issus de milieux défavorisés que pour des enfants de milieux favorisés.

Il est juste d'aider les enfants de milieux défavorisés plus que les enfants de milieux favorisés. »

J'ai également demandé aux participants de justifier leurs réponses, et d'indiquer leur genre, leur âge, leurs attitudes politiques, et leur niveau d'éducation.

Résultats

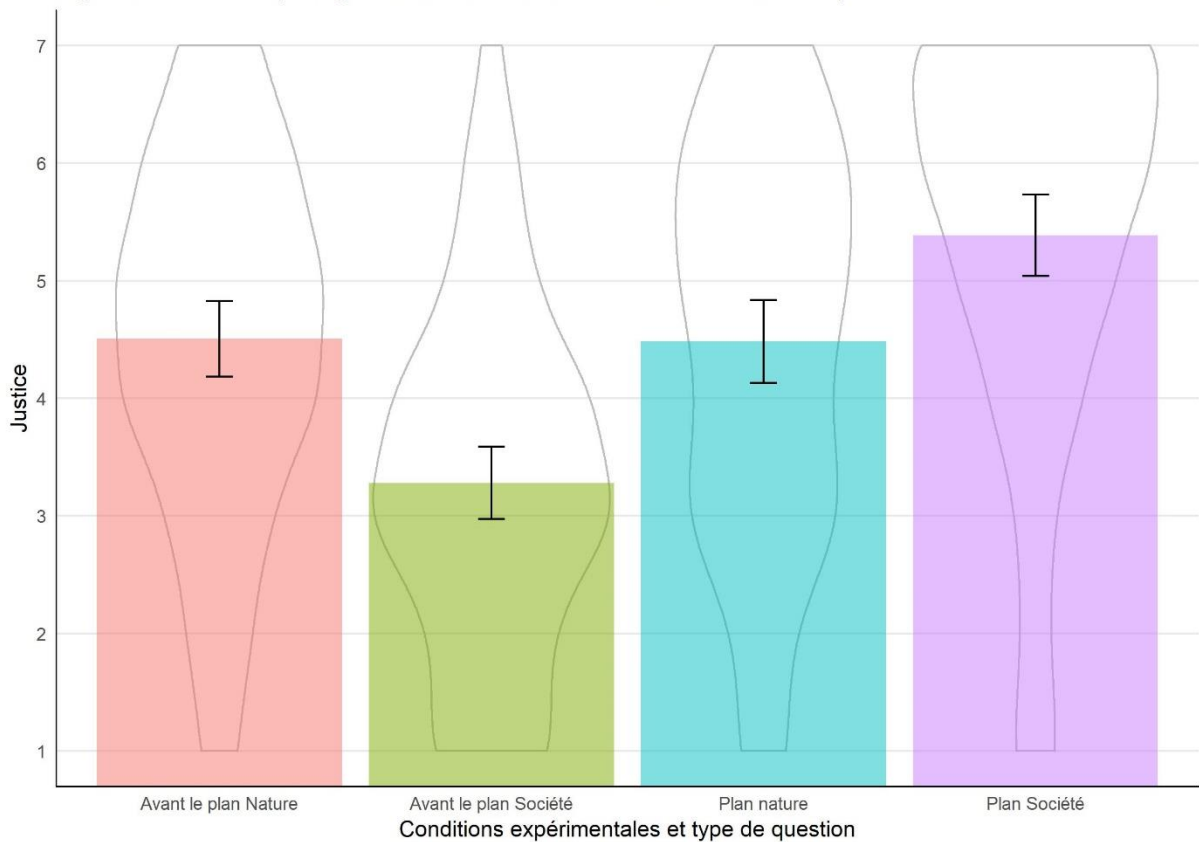
Par simplicité, dans la présentation des résultats, je me concentrerai sur l'analyse des deux questions : « Mettre en place le plan et donner des cours particuliers aux enfants de milieux défavorisés était une décision juste. » (Justice du plan) et « Komland aurait été une société juste, même si elle n'avait pas décidé de mettre en place le plan. » (Justice de la société d'avant le plan). Les deux autres questions posées donnent des résultats concordants.

Comme le lecteur peut le voir dans la figure 18, les réponses des participants ont très fortement varié entre les conditions *Différences naturelles* et *Différences sociales*. Dans le cas de la question portant sur la justice de la société avant la mise en place du plan, la réponse moyenne dans le cas de la condition *Différences sociales* est de 3,3, contre 4,5 dans le cas de la condition *Différences naturelles* (7 indiquant le plus fort accord avec la justice de la société). Cette différence est hautement significative : $M_d = 1,23$, 95% CI [0,88, 1,58], $t(92) = 6,95$, $p < ,001$. D'un point de vue qualitatif, la société où les seules différences de capacités sont d'origine naturelle est jugée comme étant plutôt juste (moyenne de 4,5), alors que la société où

les différences de capacités sont d'origine sociale est jugée comme étant injuste (moyenne de 3,3).

Une différence comparable se retrouve dans les jugements quant à la justice de la mise en place du plan. Le plan est en effet jugé comme étant beaucoup plus juste dans le cas de la condition *Différences sociales* (moyenne de 5,4) que dans le cas de la condition *Différences naturelles* (moyenne de 4,5). Cette différence est là encore hautement significative : $M_d = -0,90$, 95% CI $[-1,17, -0,63]$, $t(92) = -6,60$, $p < ,001$. Il faut toutefois noter que le plan est considéré comme étant globalement juste dans les deux conditions, même si la justice du plan dans la condition *Différences sociales* fait bien plus consensus.

Fig. 18. Justice du plan gouvernemental et de la société d'avant le plan.



Discussion

Les résultats de l'expérience 2 sont assez nets : les différences de capacités sont jugées comme étant nettement plus injustes dans le cas de différences sociales que dans le cas de différences naturelles. Les participants montrent un soutien très clair aux politiques éducatives

en faveur d'une plus grande égalité des chances pour remédier aux inégalités créées par des différences de milieu social.

Les résultats concernant une situation d'inégalité de capacités d'origine naturelle sont plus complexes. En effet, les participants ont eu tendance à juger qu'une société où les seules différences découlent de facteurs naturels est une société plutôt juste ; mais en même temps, ils ont eu aussi tendance à juger qu'un plan visant à contrer ces inégalités naturelles est un plan juste. Comment interpréter ces résultats en apparence contradictoires ? Une interprétation possible est que le fait d'aider des personnes moins talentueuses est perçu comme un acte surrogatoire, c'est-à-dire considéré comme bon en soi, mais non obligatoire. Une telle interprétation est renforcée par le fait que très peu de participants dans la condition *Différences naturelles* ont jugé que la société était très injuste avant la mise en place du plan. Des études ultérieures seraient nécessaires pour examiner dans quelle mesure une telle hypothèse est correcte.

Pour finir, il faut bien sûr mentionner que les deux expériences de ce chapitre ont été faites auprès de participants américains. Il est difficile de savoir si elles répliqueraient auprès de participants européens. Pour les raisons évoquées dans l'appendice méthodologique à la thèse, je pense que des résultats similaires seraient obtenus en Europe. Il est bien sûr impossible d'en être certain sans conduire une réplique.

Conclusion

Ce chapitre a présenté des résultats expérimentaux étudiant si les personnes jugent que des différences de mérite ne peuvent être légitimes que si tous les participants disposaient à l'origine d'exactly les mêmes possibilités. Selon les idées développées par les égalitaristes de la fortune, en effet, seules les différences découlant purement de choix individuels peuvent être considérées comme légitimes. A travers deux expériences, nous avons montré qu'une telle perspective ne correspondait pas du tout aux intuitions populaires. L'expérience 1 a notamment montré que les participants américains considèrent qu'une différence de productivité légitime des inégalités, même dans le cas où cette différence découle ultimement de facteurs vis-à-vis desquels les individus ne sont pas responsables. Plus surprenant, les participants avaient même tendance à juger que des inégalités pouvaient être justifiées, même si elles découlaient de

facteurs clairement injustes, comme des différences d'origine sociale. Il semble que la force des critères de mérite conduit les individus à passer outre des origines causales réprouvées.

L'expérience 2 a permis de montrer que, de plus, les participants ne considèrent pas que des différences de capacités d'origine naturelle constituent une forte injustice, alors même qu'ils considèrent que des différences de capacités d'origine sociale sont fortement injustes. Cette différence dans la perception des différences de capacités sociales et naturelles a été mise en avant par l'écart dans le soutien à des politiques visant à éliminer ces inégalités de productivités.

Face à ces résultats, un égalitariste de la fortune pourra parfaitement répondre qu'il ne s'agit que d'opinions biaisées, qui sont indéfendables dans le cadre d'une discussion rationnelle. Dans le chapitre suivant, je formulerai une justification générale des intuitions populaires présentées dans ce chapitre, en montrant qu'il est possible de leur fournir un appui rationnel.

Chapitre 12

L'importance de ce dont nous ne sommes pas responsables

Dans le chapitre précédent, nous avons étudié dans quelle mesure une thèse dominante dans la littérature philosophique, l'idée que des inégalités ne seraient être justifiées que si elles sont le résultat direct de choix individuels, possède une forte popularité. Nous avons vu que cette thèse ne correspond pas aux idées courantes en matière de justice sociale. Plus précisément, nous avons vu que, au moins en ce qui concerne les inégalités découlant de différences de talents naturels, une grande partie des personnes interrogées considéraient que de faibles différences de revenu pouvaient être justifiées exclusivement à partir d'une plus grande productivité. Ce chapitre sera consacré à une justification normative de ces idées courantes.

De fait, ces opinions répandues exigent une justification. Il faut au moins avouer qu'elles présentent un caractère assez énigmatique. En effet, accepter ces idées revient à défendre l'idée que, dans certains cas, une personne peut mériter de gagner plus qu'une autre, alors même que les causes de cette inégalité sont ultimement des facteurs parfaitement arbitraires d'un point de vue moral. Il est évident qu'un individu n'a pas un plus grand mérite qu'un autre simplement parce qu'il est né avec de plus fortes dispositions pour les mathématiques, ou pour le raisonnement formel. Et pourtant, les idées courantes semblent accepter que ces facteurs puissent être à l'origine d'un plus grand mérite individuel, si ces dispositions naturelles sont transformées en un travail de plus grande qualité.

Je défendrai l'idée que cette place de l'arbitraire est inévitable si l'on veut faire sens de notre vie morale. Je défendrai donc l'idée qu'il n'y a pas d'alternative à accepter l'importance de ce dont nous ne sommes pas responsables. Cette défense se fera en grande partie en reprenant les différentes justifications que nous avons proposées du mérite au cours de cette thèse. Certes, il ne s'agira pas ici de rompre complètement avec les théories philosophiques dominantes : le choix importe bien, et la responsabilité doit avoir une place essentielle dans toute théorie morale. Mais il n'en reste pas moins que nous devons reconnaître une place à des facteurs qui demeurent en dehors du choix conscient des individus concernés.

La nécessité d'accepter l'arbitraire dans la vie morale

Pour bâtir une théorie morale, nous devons accepter un certain degré de donné, de naturel, d'irresponsabilité. Sans cela, rien ne fait sens. Cette importance de l'arbitraire pourra sembler évidente au lecteur. Tout le monde sait bien sûr que nous ne sommes pas responsables de tout, et qu'il existe une grande part de hasard et d'incertitude dans la vie humaine. Cette importance de l'arbitraire a, de plus, été mise en avant de façon très forte par Bernard Williams et Thomas Nagel il y a quarante ans, quand ils ont insisté sur la part de hasard dans les cas de fortune morale²⁷⁵. Il est cependant important de rappeler cette part de l'arbitraire, pour ne pas donner l'impression que le mérite serait le seul principe moral à être affecté par des facteurs dont nous ne sommes pas responsables.

On peut mettre en avant trois grands domaines de la vie où des obligations ou des faits moraux s'imposent à l'individu sans qu'il n'y ait aucun choix de sa part. Tout d'abord, la nature de l'individu est donnée, et nous possédons donc un statut moral du simple fait de facteurs aléatoires. Deuxièmement, l'individu naît au sein d'une famille et d'une certaine communauté, et est ainsi astreint à des obligations vis-à-vis de ses semblables, sans avoir le moins du monde choisi son lieu de naissance. Enfin, troisièmement, dans le cas qui nous occupe, les talents et les capacités individuelles sont en partie étrangers à ses choix conscients. Je vais reprendre ces trois facteurs à tour de rôle.

Premièrement, le statut moral d'un individu dépend de sa nature, de ses facultés, de sa sensibilité, voire de son appartenance à une certaine espèce. Par statut moral, j'entends le degré de respect et de protection dû à un individu ou à un être vivant. Il semble relativement évident qu'il existe des inégalités dans ce domaine entre différentes espèces. Sauf à adopter un anti-spécisme extrême, un chimpanzé ou un humain posséderont un plus grand statut moral qu'un moustique. De façon beaucoup plus controversée, on pourrait aussi dire qu'un humain dispose d'un plus grand statut moral qu'un chimpanzé. Quelle que soit la manière dont on trace les frontières entre les différentes espèces animales ou végétales, il faut avouer que c'est la nature de chaque espèce ou de chaque individu qui conduit à des inégalités de statut moral. Or, la nature des individus ou des espèces est un fait dont personne n'est parfaitement responsable. Connaître exactement les sources du statut moral est complexe ; on a pu mettre en avant l'intelligence, le caractère rationnel, la sensibilité, la morale, ou encore l'appartenance à une

²⁷⁵ B. A. O. WILLIAMS et T. NAGEL, « Moral Luck », *op. cit.*

espèce rationnelle comme sources de statut moral²⁷⁶. Je ne chercherai pas ici à résoudre ces énigmes ; mais quelle que soit la position qu'on adopte, il faut reconnaître l'arbitraire à l'origine du statut possédé par chaque individu.

Deuxièmement, nous naissons tous inscrits au sein d'un tissu familial et communautaire. Bien avant d'avoir atteint l'âge de raison et avant d'avoir pu donner notre consentement, on nous a imposé une certaine identité, et nous avons bénéficié d'un ensemble d'institutions et de soins qui nous ont permis de grandir et de devenir un être autonome. On peut mettre en avant deux conséquences morales essentielles de ce statut de dépendance premier. Tout d'abord, sauf à prendre des cas graves de maltraitance ou de persécutions, nous avons tous bénéficié du soutien et de l'affection de notre famille, et nous avons profité de l'existence d'institutions politiques qui nous ont protégés. De ces facteurs naissent un ensemble d'obligations de réciprocité vis-à-vis de notre famille et de notre communauté politique. Deuxièmement, notre éducation nous a conduit à adopter une certaine identité, qui nous est imposée tout autant qu'elle est choisie. Selon un certain nombre de théories morales²⁷⁷, cette ressemblance et ce partage d'une identité commune nous conduit à posséder des obligations de soutien mutuel vis-à-vis de nos concitoyens et des personnes partageant notre identité. Là encore, nous retrouvons un fait moral fondamental qui s'impose à nous sans que nous en soyons responsables.

Troisièmement, nous défendons ici l'idée que les capacités et les talents individuels peuvent avoir une grande importance dans la vie humaine, en tant qu'ils peuvent conduire à des différences de production ou de contribution créant à leur tour des obligations morales. Même si les talents sont en eux-mêmes arbitraires d'un point de vue moral, ils jouent un rôle essentiel dans la chaîne causale à l'origine d'un plus grand mérite des individus.

Cette idée d'une importance de facteurs arbitraires n'est bien sûr pas nouvelle. Dans leur défense du mérite, des philosophes comme Michael Walzer ou David Miller ont mis en avant les limites d'une théorie qui chercherait à fonder toute obligation morale sur les choix ou les efforts individuels²⁷⁸. Plus précisément, les deux philosophes ont insisté sur le fait qu'il est impossible d'avoir une conception cohérente du sujet qui n'accorde pas une importance capitale à une nature pré-existant à tout choix ou à tout effort.

²⁷⁶ Agnieszka JAWORSKA et Julie TANNENBAUM, « The Grounds of Moral Status », in Edward N. ZALTA (dir.), *The Stanford Encyclopedia of Philosophy*, Spring 2018., Metaphysics Research Lab, Stanford University, 2018, p.

²⁷⁷ M. WALZER, *Spheres Of Justice*, op. cit. ; D. MILLER, *Strangers in our midst*, op. cit.

²⁷⁸ M. WALZER, *Spheres Of Justice*, op. cit. ; D. MILLER, *Principles of Social Justice*, op. cit. ; David MILLER, *Justice for Earthlings: Essays in Political Philosophy*, Cambridge University Press, 2013.

Cette insistance sur l'importance d'une conception du sujet qui prenne en compte des dispositions naturelles prend tout son relief dans le cadre des débats autour de l'égalité des chances. Ainsi, David Miller, dans son ouvrage récent *Justice for Earthlings*, critique les conceptions maximales de l'égalité des chances, inspirées des théories de l'égalitarisme de la fortune²⁷⁹. Cette conception maximale voudrait faire une distinction claire et nette entre les talents, les goûts et les choix individuels, afin de chercher à mettre en œuvre une société où le destin individuel dépend uniquement des choix, sans être déterminée par les talents ou les goûts personnels non-voulus. Mais, comme le souligne Miller, il est impossible de savoir ce qu'une personne aurait fait si elle disposait d'autres goûts ou d'autres prédispositions que les siennes. Miller défend l'idée qu'une théorie du mérite, comme de la responsabilité individuelle, tombe dans l'incohérence si elle ne propose pas de conception stable de la personne. Si l'on ne part pas de la nature d'un individu, avec certaines préférences et prédispositions, comme d'un donné, il n'est plus possible d'avoir une représentation claire de ce que serait le destin d'un individu.

L'idée d'égalité des chances est une institution fondamentale de nos idées de justice, et il ne s'agit pas d'y renoncer. Mais toute théorie de l'égalité des chances se fonde sur une distinction entre la personne et ses circonstances ; et il faut bien voir qu'il est impossible de construire une distinction entre la personne et les circonstances sans inclure dans la personne certaines prédispositions innées. Comme l'écrit Miller, « L'idée est de tracer une ligne entre la personne et ses circonstances, et de dire qu'une opportunité égale règne quand les circonstances sont les mêmes pour tout le monde selon certaines dimensions pertinentes. Mais la personne doit ici être comprise comme étant déjà chargée de goûts, d'un caractère, de capacités, et ainsi de suite. [...] Nous ne pouvons pas spéculer sur les choix qu'elle aurait pu faire si elle avait été formée différemment ; elle aurait alors été tout simplement une personne différente²⁸⁰. »

Construire une idée cohérente d'égalité des chances, de responsabilité, et de mérite implique donc d'accepter une part d'arbitraire dans l'origine de nos obligations et de nos droits moraux. Le mérite n'est donc pas la seule idée morale fondée sur des facteurs vis-à-vis desquels nous ne sommes pas responsables.

²⁷⁹ D. MILLER, *Justice for Earthlings*, op. cit., chap. 4-5.

²⁸⁰ *Ibid.*, p. 136.

Pourquoi les différences de talents naturels ne posent pas de problème au mérite

Nous avons vu qu'il est parfois nécessaire, pour construire une théorie morale, de faire place à ce qui n'est pas voulu, à ce dont personne n'est responsable. Mais pourquoi est-ce que les différences de talents naturels devraient avoir une telle place ? Pourquoi favoriser certaines personnes simplement parce qu'elles ont de plus grands talents ?

La réponse à cette question est qu'il ne s'agit pas de récompenser des gens simplement parce qu'ils ont de plus grands talents. Les talents des individus n'ont, fondamentalement, aucune importance d'un point de vue moral. En soi, il s'agit juste d'un fait heureux qu'une personne soit talentueuse. Par contre, ce qui a du sens d'un point de vue moral, c'est le fait que ces talents peuvent conduire à un travail de meilleure qualité, ou à une plus grande productivité. Si une personne talentueuse met ses capacités à bon escient, alors il est important de la récompenser, pour toutes les raisons que l'on a détaillées dans les chapitres 6 à 9. En effet, cela s'accorde avec nos intuitions ; cela sert l'efficacité collective ; cela crée une compatibilité entre les intérêts personnels et les buts collectifs ; enfin, cela communique à la personne l'approbation de ce qu'elle a fait, et sert ainsi à constituer une communauté de valeurs. Ces fonctions existent toujours dans le cas d'une productivité supérieure ultimement causée par des talents individuels plus importants. Les justifications que nous avons fournies du mérite constituent en même temps une justification de l'importance des talents naturels, si ceux-ci sont mobilisés vers une plus grande contribution.

Talents naturels et capacités sociales : quelle différence ?

Cependant, on peut considérer que les arguments qui viennent d'être cités constituent aussi une justification des inégalités découlant ultimement de facteurs clairement injustes. Par exemple, reprenons le cas de deux personnes, possédant des talents naturels similaires, mais qui ont des capacités productives à l'âge adulte très différentes, parce que l'une a grandi dans un milieu très favorisé, et l'autre dans un milieu très défavorisé.

Dans ce cas, si la personne ayant grandi dans un milieu favorisé fait un travail de meilleure qualité, les arguments que j'ai mobilisés ci-dessus semblent défendre le fait qu'elle mérite de

gagner plus que l'autre. En effet, trois des quatre justifications (incitations, accord entre l'intérêt personnel et l'intérêt collectif, communication) disent clairement qu'il faut récompenser la personne ayant fait un travail de bonne qualité. La quatrième justification (les intuitions courantes) donne une réponse plus complexe puisque, comme nous l'avons vu au chapitre précédent, les participants de nos expériences étaient partagés vis-à-vis de la justice d'inégalités de capacités dérivant ultimement de facteurs injustes. Cependant, une partie substantielle des gens disait qu'il était juste de récompenser davantage la personne ayant fait un travail de meilleure qualité. S'agit-il d'une *reductio ad absurdum* des justifications que nous avons proposées, puisqu'elles semblent justifier des inégalités ayant leur origine dans une différence de milieu social ?

Certes non. Deux facteurs normatifs sont à l'œuvre ici, qui vont dans deux directions opposées. Tout d'abord, les différentes justifications du mérite nous disent qu'il faut récompenser le travail bien fait. Mais, à l'opposé, nous avons l'intuition qu'une injustice a été faite, et qu'il faut compenser celle-ci. Les inégalités d'origine sociale ont en effet un tout autre statut normatif que les inégalités de talent. Dans le cas des inégalités sociales, la société et l'État avaient une obligation déontologique de traiter tous les citoyens de la même manière. Le fait que certaines personnes ont bénéficié d'avantages personnels constitue alors une violation des devoirs que la société a vis-à-vis de chaque individu. Dans ce cadre, il est possible de comprendre pourquoi nous avons des intuitions partagées dans le cas des inégalités d'origine sociale, mais pas vis-à-vis des inégalités d'origine naturelle.

Que faire dans une société injuste ? Mérite compétitif et mérite non-compétitif

A partir du moment où l'on accepte qu'il y a une différence de statut entre inégalités naturelles et sociales, il reste à déterminer comment il faut prendre en compte les différences de mérite dans le cas d'une société injuste. Étant donné que les justifications que nous avons mises en avant pour défendre le mérite tiennent toujours dans ce cas, il semble clair qu'un meilleur travail implique toujours un mérite supérieur. Pour autant, il semble difficile de laisser de côté l'origine injuste des différences de productivité, si celles-ci sont principalement dues à l'existence de milieux sociaux favorisés et défavorisés. Comment concilier la récompense du travail et la compensation due aux injustices sociales ?

Tout d'abord, il faut mettre en avant l'existence de deux types de contextes où un individu peut effectuer une contribution. Ces deux contextes sont les contextes compétitifs et non-compétitifs.

Une circonstance de contribution est non-compétitive si les individus ne possèdent pas un droit fondamental au fait de contribuer. A la suite de Miller, je défends l'idée que des circonstances injustes n'affectent le mérite des individus que dans le cas de circonstances compétitives²⁸¹.

Pour illustrer ce qu'est une circonstance non-compétitive, reprenons le cas où une personne se noie. Dans ce cas, la personne qui l'a sauvée mérite pleinement les remerciements. Étant donné qu'il s'agit d'un cas d'urgence vitale, personne ne possède un droit fondamental à participer au sauvetage de la personne. Celui qui est en moyen de sauver la personne doit le faire, et c'est tout. Il n'y a pas de compétition devant être organisée entre les différentes personnes pouvant éventuellement sauver la personne.

Malheureusement, dans la plupart des cas intéressants en matière de justice sociale, nous sommes confrontés à des cas de circonstances compétitives. En effet, une des sources les plus fréquentes de mérite dans nos sociétés modernes passe par le travail. Or, on peut légitimement considérer que la possibilité d'exercer un emploi constitue un droit fondamental. Ce droit fondamental découle du fait que l'État a le devoir d'assurer à tous le plein développement de leurs capacités, et du fait qu'accomplir un travail constitue un facteur majeur de l'épanouissement humain. En conséquence, nous disposons d'un droit de regard collectif sur la manière dont l'attribution d'emplois se fait²⁸². De ce fait, le travail constitue un contexte compétitif d'acquisition du mérite.

La place du mérite : estimer le mérite contrefactuel

Il nous faut déterminer comment évaluer le mérite dans des circonstances injustes. Dans cette section, je vais défendre l'idée que l'évaluation du mérite des personnes dans un monde injuste implique d'effectuer une estimation du mérite que les personnes auraient obtenu dans

²⁸¹ D. MILLER, *Principles of Social Justice*, op. cit., p. 145.

²⁸² Ce droit de regard collectif est ainsi particulièrement marqué en ce qui concerne les jugements de discrimination.

une société parfaitement juste. Ce mérite contrefactuel doit alors servir de moyen à évaluer le mérite des individus dans le monde réel.

Si l'on compare deux personnes, l'une ayant grandi dans un environnement défavorable, et l'autre dans un environnement favorable, il nous est difficile d'imaginer ce qui aurait pu se passer si les deux personnes avaient grandi dans le même type d'environnement. A strictement parler, on ne sait pas ce qu'il serait advenu si quelqu'un n'avait pas subi un désavantage social. Cependant, il est toujours possible de faire une estimation de ce qu'il aurait pu se passer dans une société juste. Il faut donc prendre en compte deux types d'actions méritoires : d'une part, les actions méritoires qui ont été faites dans le monde réel ; d'autre part, les actions méritoires qui auraient été faites si les deux individus avaient bénéficié d'un environnement également favorable.

Résoudre le problème du mérite dans un monde injuste dépend de nos possibilités d'estimer le mérite dans un monde idéal. Commençons par un exemple extrême pour montrer qu'un choix informé dans l'estimation du mérite idéal n'est pas impossible. Si nous comparons le mérite d'Einstein avec celui d'un individu qui a grandi dans des circonstances beaucoup moins favorables, nous pouvons être presque certains que l'individu moins favorisé ne serait pas devenu un des plus grands scientifiques de tous les temps s'il était né dans une famille bourgeoise ou intellectuelle. De fait, quasiment personne ne devient Einstein, quelle que soit la famille dans laquelle on naît. Le mérite d'Einstein reste donc presque entier, même dans une société injuste.

Dans des cas moins extrêmes que celui d'Einstein, le doute est davantage permis. Imaginons qu'un enfant de bonne famille devienne ingénieur, tandis qu'une personne étant née dans une famille beaucoup moins favorisée obtient un travail plus facile à faire, et moins bien payé. Est-il possible de dire que l'ingénieur mérite de gagner plus que l'autre ? Il n'est pas certain que, dans un monde parfaitement juste, la personne qui est devenue ingénieur aurait fait des contributions plus méritoires que l'autre. Cependant, étant donné qu'ingénieur est un emploi difficile, nous pouvons quand même dire qu'il est probable que cela aurait été le cas ; probable, mais non certain. La conséquence de cette incertitude est d'opérer une régression vers la moyenne : notre estimation du mérite de l'ingénieur dans une société contrefactuelle est que son mérite est supérieur à la moyenne, mais moindre que si l'on savait avec certitude qu'il serait également devenu ingénieur dans une société parfaitement juste.

La conséquence normative de ces considérations est que les différences de contributions perdent partiellement leur valeur légitimatrice dans une société injuste. Dans le cas de personnes ayant fait des contributions importantes, il reste probable qu'elles auraient agi positivement

même dans un monde où elles n'auraient pas été ainsi favorisées ; comme nous n'avons aucune certitude à ce sujet, leur estimation de mérite diminue, sans toutefois revenir à une estimation de mérite parfaitement égale à la moyenne. La conséquence est que les inégalités justifiées par des différences de capacités contributives sont plus faibles dans une société injuste que dans une société juste. En d'autres termes, l'absence d'égalité des chances rend les inégalités de revenu moins légitimes. Une telle conclusion est en accord avec les intuitions courantes relatives à ce sujet que nous avons étudiées dans le chapitre 11.

Le mérite peut donc garder du sens même en cas d'absence d'égalité des chances, mais il en est diminué. Il ne prend un sens maximal que dans un contexte où la différence de production entre deux personnes peut s'expliquer exclusivement à partir de choix, de talents et d'efforts.

Le rôle suprême du choix

Avant de conclure, il faut revenir sur l'importance du choix et de la responsabilité individuelle. J'ai mis en avant dans ce chapitre la valeur de facteurs dont nous ne sommes pas responsables. Cela ne veut bien sûr pas dire que la responsabilité n'est pas, en général, plus importante pour légitimer de possibles inégalités que des différences de talents naturels. De fait, il me semble clair qu'une inégalité sera plus juste si elle est due à des choix qu'à des facteurs pour lesquels les individus ne sont pas responsables. En ce sens, l'égalitarisme de la fortune avait raison.

Cette importance du choix découle de nombreuses raisons. Thomas Scanlon a notamment détaillé trois raisons expliquant la valeur du choix²⁸³. Tout d'abord, laisser le choix aux individus possède une valeur instrumentale : les individus savent en général mieux que d'autres ce qui est bon pour eux. On peut ajouter à cela que, en général, laisser les individus subir les conséquences de leurs choix crée un système d'incitations favorables, en tant que les individus sont dissuadés de faire peser sur d'autres des externalités négatives. Deuxièmement, le choix possède dans certaines circonstances une valeur représentative. Mon choix représente mes goûts et ma personnalité, et il est bon que ma vie soit le reflet de ma personne. Ainsi, il est préférable que des individus choisissent leur emploi, plutôt qu'il leur soit imposé de l'extérieur. Enfin, le choix possède une valeur symbolique : choisir à la place des individus voudrait dire qu'on ne

²⁸³ T. M. SCANLON, « The Significance of Choice », *op. cit.* ; T. M. SCANLON, *What We Owe to Each Other*, Cambridge, Mass., Belknap Press, 2000, chap. 6.

les considère pas capables de décider pour eux-mêmes. Laisser les individus libres de leurs choix est ainsi un symbole de l'égalité entre citoyens. Ces trois valeurs, instrumentales, représentatives, et symboliques, sont cruciales pour comprendre l'importance de la responsabilité dans les intuitions courantes relatives à la justification des inégalités.

Malgré l'existence de toutes ces raisons, il n'en serait pas moins absurde d'accorder un caractère sacré au choix. Nous pouvons avoir de bonnes raisons de limiter le choix ou de prendre soin de quelqu'un même s'il a pris de mauvaises décisions. L'idée de paternalisme peut être gênante, mais il est souvent préférable de forcer les individus à agir pour leur propre bien. Un exemple classique correspond à la mise en place d'un système de cotisations forcées pour les retraites, qui limite le biais court-termiste inhérent à l'homme. De la même manière, nous pouvons valoriser l'impact des talents naturels, sans qu'ils aient été eux-mêmes choisis.

Conclusion

Dans ce chapitre, nous avons défendu que la tentation de fonder la moralité toute entière sur la base de l'idée de responsabilité doit être refusée. De nombreux facteurs dont nous ne sommes pas responsables, comme notre nature, notre lieu et famille de naissance, nos talents, peuvent légitimement influencer les devoirs moraux que nous avons les uns vis-à-vis des autres. Les différentes justifications du mérite que nous avons détaillées tout au long de cette thèse permettent bien de dire qu'une personne peut mériter davantage qu'une autre à partir du moment où elle a effectué un travail de meilleure qualité, quelle que soit la raison ultime à l'origine de ce travail supérieur.

Les cas de différences de productivité dues à des origines sociales inégales créent, cependant, un cas particulier. En effet, ces situations font entrer en conflit deux principes opposés : d'une part l'idée qu'un travail supérieur mérite un revenu supérieur, et d'autre part, le fait que la société a le devoir de réparer l'injustice qui a été causée par les différences de milieu social. Dans ce cas, nous avons défendu l'idée qu'il faut prendre en compte deux types de contributions méritoires : d'une part la contribution réelle, et d'autre part la contribution contrefactuelle qui aurait eu lieu dans une société juste. Une telle situation diminue l'étendue des inégalités qui peuvent être justifiées par des différences de contribution dans le monde réel.

Conclusion

Défense du mérite et théorie pluraliste de la justice

Cette thèse a procédé en trois temps pour effectuer une théorie générale et une défense du mérite. Dans un premier moment, nous nous sommes efforcés de proposer une définition suffisamment précise de l'idée de « mérite ». Le principal but de ce travail de définition a été de défendre l'idée qu'une conception intéressante du mérite ne doit pas être réductible à d'autres idées morales, comme l'égalité, ou comme le devoir de satisfaire les attentes légitimes des individus. Nous avons également défendu l'idée que le mérite ne devait pas être réduit à une idée purement morale, comme la récompense de la vertu. En effet, il est important de garder une conception suffisamment large du mérite, qui puisse correspondre aux usages quotidiens, comme l'idée que tout travail mérite salaire. Cela nous a permis de montrer qu'une théorie intéressante et politiquement pertinente du mérite devait avant tout expliquer les jugements qu'un individu doit être récompensé ou puni pour une bonne ou mauvaise action, et ce, pour des raisons en partie intrinsèques, et non purement instrumentales.

Dans cette première partie, nous avons également montré qu'aucun des critères habituels du mérite, la contribution ou les efforts, ne peuvent être considérés comme des critères plausibles, et ce, quelle que soit la justification que l'on cherche à apporter à l'idée de mérite. Ces deux critères conduisent en effet à des situations burlesques d'un point de vue normatif, comme l'idée qu'il faudrait récompenser des individus pour des efforts parfaitement inutiles, ou pour une contribution produite par pur hasard. Nous avons alors montré qu'un critère légèrement modifié, à savoir l'espérance de contribution, permet de résoudre les problèmes habituellement associés à la détermination d'un critère du mérite. Cependant, ce nouveau critère est souvent très délicat à mettre en œuvre, et il est important de trouver des moyens approximatifs de mesurer l'espérance de contribution. Cette nécessité de trouver des critères approximatifs explique pourquoi la contribution et les efforts peuvent souvent servir d'heuristiques dans la détermination du mérite.

Dans un deuxième temps, nous avons procédé à une justification générale de l'idée de mérite. Cette défense s'est basée sur quatre justifications principales. Tout d'abord, une première justification s'est basée sur le fait que nous avons des intuitions morales fortes en faveur de l'idée de récompenser les individus selon leur mérite. L'idée que nous disposons d'intuitions fortes en faveur du mérite s'est fondée sur le fait que le mérite est une idée reposant sur un consensus largement partagé, aussi bien au sein de nos pays qu'à l'international.

Ensuite, nous avons également mis en avant une justification instrumentale du mérite. En effet, la récompense du mérite favorise en général l'existence d'un système incitatif conduisant à une plus grande productivité générale. Certes, nos intuitions relatives au mérite sont souvent déconnectées de la recherche de la plus grande efficacité, comme quand nous considérons que nous devons récompenser des personnes particulièrement altruistes, même si celles-ci auraient de toute façon agi de la même façon avec ou sans récompense. Malgré tout, en général, la récompense du mérite sert l'efficacité. Une justification instrumentale, si elle serait insuffisante pour justifier à elle seule l'existence du mérite, peut servir à nous rassurer quant au fait que la récompense du mérite ne dessert en général pas la recherche de l'intérêt général.

Pour une troisième justification, nous avons mis en avant le fait que la récompense du mérite sert à préserver l'existence d'une congruence et d'un ordre entre nos différents systèmes de valeurs. Plus précisément, la récompense du mérite assure le fait qu'il n'existe pas de divorce entre l'intérêt individuel et l'intérêt collectif, en tant que les actions bonnes pour le groupe deviennent, par le biais de la récompense du mérite, aussi bonnes pour l'individu.

Enfin, nous avons défendu que la récompense du mérite sert à assurer une communauté de valeurs, en tant que la récompense du mérite permet de communiquer à l'individu méritant que ce qu'il a fait est approuvé par la communauté. Nous avons montré par des études expérimentales que cette valeur communicative du mérite est ancrée dans la psychologie ordinaire de la peine.

Une fois ces justifications du mérite assurées, nous avons tenté de répondre à une objection classique à l'encontre du mérite, à savoir, le fait que la récompense du mérite ne ferait pas de sens d'un point de vue moral parce qu'elle exigerait une conception trop exigeante de la responsabilité. Nous avons alors mis en avant le fait que, pour faire sens, le mérite ne demandait qu'une conception extrêmement limitée de la responsabilité, à savoir l'idée d'un simple contrôle proximal sur nos actions. Nous avons, de plus, montré que le fait que nous ne sommes pas responsables de nos talents individuels ne constituait en rien un problème pour une conception du mérite, qui doit pouvoir accepter comme point de départ des situations qui ne relèvent de la responsabilité de personne.

En soi, le mérite possède donc une valeur, et il ne présuppose pas une préconception de l'idée de responsabilité si exigeante que le mérite deviendrait impossible. Il s'agit, en d'autres mots, d'une idée valable et globalement réaliste.

Il nous reste maintenant à esquisser la résolution d'un dernier type de difficultés qui se posent pour le mérite. J'ai jusqu'à présent bâti une justification générale du mérite presque sans prendre en compte la relation du mérite à d'autres valeurs, comme l'égalité. La théorie du mérite s'est faite sans construire une théorie globale de la justice sociale. Ce parti pris se fonde sur une idée simple : tant que l'on n'adopte pas une position moniste en matière de justice sociale, c'est-à-dire tant que l'on ne considère pas que la justice sociale doit se fonder sur un principe unique, il est important d'étudier chaque principe potentiel en son nom propre, avant de voir comment ces principes se combinent ensuite les uns avec les autres. Dans cette conclusion, je vais détailler davantage une telle perspective pluraliste.

Tirer les conclusions de l'importance du mérite : la valeur du pluralisme

Il n'existe pas de raison de réduire les différents principes de justice à une seule idée. Nous disposons d'une pluralité d'intuitions relatives à la justice, et il me semble relativement vain de considérer qu'il est possible réduire qu'il est possible, par exemple de réduire notre attachement vis-à-vis du mérite à un type d'égalité. Ce pluralisme moral devient d'autant plus nécessaire si l'on considère l'ensemble de nos idées morales, et non plus les seules considérations de justice. Même si la maximisation du bien-être collectif n'est pas forcément une question de justice, il s'agit à peu près à coup sûr d'une question morale. De la même manière, la relation entre liberté et justice est complexe, certains auteurs considérant que la promotion de la liberté individuelle est un devoir de justice, d'autres n'intégrant pas ces questions à leur théorie de la justice²⁸⁴. Il me semble ici aussi difficile de dire que l'idée de liberté et d'égalité correspondent à une seule

²⁸⁴ Pour des théories intégrant la liberté à leur théorie de la justice, on peut voir les théories élaborées par John Rawls ou Robert Nozick ; pour des théories le refusant, on peut voir par exemple la théorie pluraliste proposée par David Miller. J'ai personnellement l'intuition qu'il s'agit de deux questions séparées, et que certaines théories de la justice commettent une erreur de catégorie en considérant que la promotion des libertés individuelles est en soi une question de justice. R. NOZICK, *Anarchy, State and Utopia*, op. cit. ; J. RAWLS, *Théorie de la justice*, op. cit. ; D. MILLER, *Principles of Social Justice*, op. cit.

et même idée, même si l'on considère qu'il n'existe aucune incompatibilité entre ces deux notions.

Bâtir une théorie pluraliste de la justice dépasse notre propos ici ; je renvoie aux théories défendues notamment par David Miller et Kok-Chor Tan, qui me semblent être les théories pluralistes les plus plausibles et les mieux élaborées à l'heure actuelle²⁸⁵. David Miller a construit une théorie sophistiquée, construisant la justice à partir de trois principes de justice distincts : la satisfaction des besoins, le mérite, et l'égalité. Kok-Chor Tan a de son côté construit une théorie accordant une priorité à la satisfaction des besoins des citoyens, et défendant une théorie égalitariste de la fortune pour toute distribution supérieure à la satisfaction des besoins fondamentaux. Miller a tout particulièrement insisté sur l'accord général entre la théorie qu'il défend et les intuitions courantes à ce sujet²⁸⁶.

Si l'on adopte une telle perspective pluraliste sur la justice, alors l'intégration entre le mérite et les autres valeurs que l'on pourrait défendre ne pose pas nécessairement de problèmes insolubles. Si l'on considère à la fois que le mérite et l'égalité possèdent de la valeur, alors la question n'est pas de savoir quel principe de justice éliminer. La question est, plutôt, de chercher à construire une société qui permette au maximum de satisfaire les demandes potentiellement conflictuelles de ces deux idées. Il s'agit, fondamentalement, d'un art du compromis. Bien sûr, cette recherche du compromis pose des problèmes philosophiques particuliers ; il s'agit, cependant, d'une question philosophiquement moins dramatique que celle de savoir dans quelle mesure le mérite possède de la valeur en tout premier lieu.

Construire une théorie pluraliste de la justice implique de repérer les tensions potentielles entre le mérite et d'autres idées morales. La tradition philosophique a souligné deux sources de tensions principales. Tout d'abord, il existe bien sûr un conflit potentiel avec l'idée d'égalité. Ensuite, trois des théoriciens majeurs du mérite et de la justice sociale, Henry Sidgwick, Friedrich Hayek, et John Rawls, ont souligné que le mérite est incompatible avec le fonctionnement des économies modernes, basées sur la loi du marché. Si l'on considère que le marché sert deux buts moraux fondamentaux, à savoir l'efficacité et la liberté individuelle²⁸⁷, alors on aboutit bien à une incompatibilité entre le mérite et d'autres valeurs importantes. La

²⁸⁵ D. MILLER, *Principles of Social Justice*, op. cit. ; K.-C. TAN, *Justice, Institutions, and Luck*, op. cit. Une théorie pluraliste classique est bien sûr celle défendue par Michael Walzer. Une autre théorie pluraliste récente accordant une grande importance à l'égalité est celle défendue par David Schmidtz. M. WALZER, *Spheres Of Justice*, op. cit. ; D. SCHMIDTZ, *The Elements of Justice*, op. cit.

²⁸⁶ D. MILLER, *Principles of Social Justice*, op. cit.

²⁸⁷ Sur ces deux fonctions morales du marché, voir les textes classiques d'Amartya Sen, et notamment *Development as freedom*, Oxford, Univ. Press, 2001.

section suivante abordera la compatibilité entre le mérite et le fonctionnement d'une économie de marché. Je conclurai enfin par un bref examen du lien entre le mérite et l'égalité.

Le conflit entre la récompense du mérite et l'existence d'une société de marché

Une critique classique du mérite, que l'on retrouve aussi bien chez Sidgwick, Hayek et Rawls, consiste à souligner que le marché est insensible au mérite des individus. En conséquence, la récompense du mérite ne saurait être qu'une chimère dans nos sociétés modernes.

L'argument des trois auteurs est le suivant. Selon eux, le mérite est supposé être une donnée stable et propre aux individus. Le mérite ne doit donc pas dépendre des fluctuations de l'offre et de la demande. Or, dans les économies de marché, les salaires dépendent essentiellement de la loi de l'offre et de la demande. En conséquence, le mérite ne peut pas être mis en place dans nos sociétés modernes.

Il existe deux manières différentes de répondre à cette objection. La première manière, choisie par David Miller, consiste à souligner que le marché récompense le fait de satisfaire les désirs d'autres individus. En ce sens, le marché consiste en grande partie à récompenser le mérite des individus. Une telle thèse est plausible. Dans un marché libre, il existera à peu près à coup sûr une corrélation positive entre le mérite des individus et leur salaire, en ce qu'effectuer davantage d'efforts conduira en moyenne à un meilleur salaire, de même que le fait d'avoir davantage de talents, si ces talents sont mis en œuvre à bon escient. En ce sens, il semble bien que la récompense du mérite constitue une justification supplémentaire de l'existence d'un marché économique.

Cependant, il semble difficile de dire que le marché récompensera systématiquement le mérite des individus, pour un certain nombre de raisons bien connues des économistes. Tout d'abord, il faut admettre que les objections posées par Sidgwick, Hayek et Rawls ont une certaine force. Selon les critères mis en avant dans notre chapitre 2, nous considérons que le critère du mérite est l'espérance de contribution d'une action productive. Mais il n'y aura jamais une parfaite concordance entre ce que le travail d'un individu devrait apporter aux autres, et ce que cette personne obtiendra finalement sur le marché du travail. Si l'on imagine un individu produisant une certaine quantité de blé, par exemple, alors il s'agit d'une activité qui constituera

en général une forte contribution au bien-être collectif. Mais si l'on imagine que, pour des raisons conjoncturelles et imprévisibles, le prix du blé s'effondre une année, alors la personne recevra un revenu très faible cette année-là, sans que cela corresponde à une moindre espérance de contribution de l'individu. Dans ce cadre, même si le revenu qu'un individu peut obtenir sur le marché tendra à être corrélé avec ce qu'il produit, cette corrélation sera toujours imparfaite²⁸⁸.

De façon plus générale, il est évident que les salaires obtenus par les individus sur le marché du travail ne correspondent pas forcément avec ce que nous entendons par le mérite des individus. Si l'on pense au fait qu'il existe une très forte augmentation du chômage au sein de la société, alors les entreprises disposeront d'un pouvoir de pression supplémentaire pour faire baisser les salaires des individus. Dans ce cadre, du fait de l'augmentation de l'offre de travail, à demande constante, le prix des travailleurs baissera forcément, si l'on fait confiance à de pures règles de marché. Il semble qu'il existe un conflit très net entre les mécanismes de marché et la récompense du mérite.

Heureusement, il n'existe pas de raison de penser que, s'il existe un conflit entre ce que nos intuitions de mérite dicteraient et ce qu'une analyse purement économique en termes d'offre et de demande exigerait, il y aura toujours une défaite de nos intuitions méritocratiques. Afin de critiquer le mérite, Sidgwick, comme Hayek, comme Rawls, semblent présupposer que le marché du travail fonctionne sur la pure base de raisons économiques, sans être du tout traversé par des considérations morales ou de valeurs. Une telle idée a sérieusement été remise en cause au cours des quarante dernières années. Aussi bien les entreprises, que les syndicats, que les travailleurs eux-mêmes, accordent de l'importance au mérite, et conduisent à un fonctionnement de l'économie qui est très différent de celui prédit par des pures considérations d'offre et de demande. George Akerlof a ainsi reçu le prix Nobel d'économie en grande partie pour ses travaux insistant sur l'importance des sentiments de justice pour prédire le

²⁸⁸ Il serait possible de mettre en avant beaucoup d'autres facteurs pour rendre compte du fait que le marché ne récompense pas toujours le mérite des individus. Je me contenterai de mettre en avant un fait particulièrement important. L'économie de marché réelle ne ressemble pas vraiment à un modèle de concurrence pure et parfaite. Il existe de nombreuses rentes de situation, où certaines entreprises peuvent profiter d'une situation de monopole dans le cadre de secteurs où il est difficile de concurrencer les entreprises établies, par exemple à cause de l'existence d'économies d'échelles ou de dépendance au sentier qui favorisent les entreprises en place. Dans ce cadre, certains groupes, comme les géants des nouvelles technologies aujourd'hui, disposeront systématiquement d'un revenu gigantesque, sans qu'il soit évident que ces revenus extravagants soient parfaitement mérités.

fonctionnement du marché²⁸⁹. Je prendrai trois exemples indiquant l'importance de nos considérations de justice²⁹⁰.

Tout d'abord, les salaires des individus ne sont pas uniquement le résultat d'une négociation entre le patron et son salarié, mais relèvent d'un encadrement collectif notamment visible par le biais des grilles de salaire. En France, notamment à partir de l'action développée lors de la Première Guerre Mondiale et lors du Front populaire, de nombreuses négociations entre les syndicats et le patronat ont eu lieu pour déterminer quelle grille de salaire devait s'appliquer à l'échelle de l'entreprise. Cette détermination des échelles salariales par la négociation a perduré jusqu'à aujourd'hui, et peut prendre en compte des critères comme l'expérience, la qualification ou la pénibilité. Au sein de l'entreprise en tout cas, la détermination du mérite des différents individus est une pratique importante. On peut critiquer la façon dont sont mises en place ces grilles de salaire, notamment le fait qu'elles séparent les cadres des autres employés ; mais il est difficile de nier leur impact réel sur l'économie²⁹¹.

Pour un deuxième exemple, on peut mettre en avant le fait que les collectivités peuvent aller à l'encontre de leur intérêt économique pour respecter des considérations de mérite. Un exemple amusant est fourni par le cas du salaire des professeurs dans les universités américaines. Vue la richesse des débouchés qui attendent les scientifiques, aussi bien dans le public que dans le privé, comparée à la pauvreté des débouchés possibles pour les littéraires, une université respectant la loi de l'offre et de la demande paierait ses professeurs littéraires beaucoup moins que ses professeurs scientifiques. Or, ce n'est pas le cas, puisque les professeurs scientifiques sont à peine plus payés que les professeurs littéraires, ce qui occasionne des difficultés notables de recrutement pour les postes scientifiques. Contrairement à toute logique économique, les universités préfèrent respecter le travail égal des scientifiques et des littéraires²⁹².

Les entreprises et les syndicats accordent donc de l'importance aux considérations de mérite. Mais c'est aussi le cas des salariés individuels. De nombreuses expériences ont étudié les effets d'un salaire faible (comparé au salaire des travailleurs remplissant la même fonction) sur la motivation d'un individu. Les résultats sont clairs : un salarié qui s'estime trop peu payé par

²⁸⁹ George Arthur AKERLOF, Robert James SHILLER et Corinne FAURE-GEORS, *Les esprits animaux: comment nos comportements irrationnels gouvernent l'économie*, 2018 ; George A. AKERLOF et Janet L. YELLEN, « Fairness and Unemployment », *The American Economic Review*, 1988, vol. 78, n° 2, p. 44-49.

²⁹⁰ Les paragraphes suivants reprennent en partie la discussion plus détaillée de ces questions que j'ai effectuée dans mon mémoire de maîtrise, écrit sous la direction de Sophie Guérard de Latour. Aurélien ALLARD, « Mérite et travail ».

²⁹¹ Jean SAGLIO, « Hiérarchies salariales et négociations de classifications - France, 1900-1950 », *Travail et Emploi*, 1986, N°27 ; Jean SAGLIO, « Négociation de classifications et régulation salariale dans le système français de relations professionnelles », *Travail et Emploi*, 1988, N°38.

²⁹² D. KAHNEMAN, J. L. KNETSCH et R. THALER, « Fairness as a Constraint on Profit Seeking », *op. cit.*

rapport à ses mérites diminuera fortement ses efforts, pour s'adapter au salaire qui lui est effectivement payé. De telles réactions étant connues des employeurs, elles peuvent tendre à dissuader l'employeur de baisser les salaires des employés²⁹³.

Il ne s'agit pas de faire du monde du travail un univers dans lequel le mérite est toujours respecté et défendu. Ce n'est pas le cas ; les lois économiques d'offre et de demande y jouent bien un rôle important, de même que les relations de domination ou de contrôle, qui contraignent certains salariés à accepter des heures supplémentaires non payées ou des emplois très mal rémunérés. Cependant, le travail est un milieu mixte, où la défaite de l'idée de mérite n'est jamais gagnée d'avance. Il n'y a pas de raison de rejeter un principe de justice simplement parce qu'il n'est pas toujours strictement appliqué dans la vie réelle. C'est précisément le but d'une idée morale de pouvoir nous guider dans la critique du monde actuel, et notamment d'un marché qui accorderait trop peu d'importance aux efforts ou aux contributions des individus.

Réconcilier le mérite et l'égalité

Nous avons donc vu que l'opposition possible entre mérite et économie de marché ne conduisait pas nécessairement à l'élimination de l'idée de mérite. Il faut maintenant considérer une opposition qui peut sembler à première vue beaucoup plus grave, à savoir la possible contradiction entre égalité et récompense du mérite. Je défendrai ici l'idée que l'opposition entre mérite et d'égalité a été très fortement exagérée, et qu'il existe un certain nombre de raisons de penser que les deux idées sont relativement compatibles.

Le mérite et l'égalité ont traditionnellement été pensés comme opposés. Le mérite a de nombreuses fois été représenté comme étant une source d'inégalité et comme étant une justification des inégalités. Défendre l'égalité reviendrait à remettre en cause le mérite. Il ne s'agit pas ici de remettre en cause l'idée qu'il existe une certaine tension entre mérite et égalité. Sauf à faire l'hypothèse que tous les membres d'une société possèdent un mérite parfaitement égal, toute distribution sensible au mérite des individus correspondra fatalement à une distribution au moins partiellement inégalitaire.

Cependant, un tel conflit entre mérite et égalité est probablement moins gênant, ou moins décisif, qu'il ne peut le sembler à premier abord. En effet, presque personne, ni parmi les

²⁹³ G. A. AKERLOF et J. L. YELLEN, « Fairness and Unemployment », *op. cit.*

philosophes, ni parmi les mouvements politiques, ni parmi les citoyens, ne considère qu'une situation d'égalité parfaite est une situation viable. Toutes les théories majeures de la justice, formulées au cours de ces cinquante dernières années, tolèrent l'existence d'un minimum d'inégalités, que ce soit au nom de la nécessité de fournir des incitations ou au nom de la responsabilité individuelle. Par ailleurs, il semble peu probable que nous arrivions à une situation d'égalité parfaite dans des sociétés complexes et plurielles actuelles.

A partir du moment où l'on accepte la probable existence d'inégalités économiques au sein de la société, deux questions peuvent se poser : premièrement, quelle magnitude d'inégalités accepterons-nous ? Deuxièmement, qu'est ce qui explique ou justifie l'existence de telles inégalités ? Les deux types de questions sont assez largement indépendantes. Par rapport à la première question, on peut imaginer un type de société où l'écart des revenus ou des richesses entre les plus pauvres et les plus riches est très faible, ou bien une société où l'écart des revenus entre plus pauvres et plus riches est gigantesque. Par rapport à la deuxième question, on peut imaginer que les différences de richesse s'expliquent essentiellement par le mérite, ou bien par le hasard, l'héritage, les droits de propriété, ou la corruption. Il n'y a pas de lien nécessaire entre les deux questions.

Prenons le cas de l'image stéréotypique de quatre sociétés : la Suède, les États-Unis, l'Union Soviétique, la Chine actuelle. La Suède pourrait représenter un cas où l'étendue des inégalités est faible, et où les inégalités sont basées sur des raisons méritocratiques, du fait de la qualité du système scolaire et de la promotion de l'égalité des chances. L'image stéréotypée des États-Unis pourrait représenter un cas où l'étendue des inégalités est très large, mais ces inégalités sont expliquées par des raisons de mérite, liées au fait qu'un *self-made man* peut s'imposer s'il en a les capacités²⁹⁴.

Nous avons donc deux questions différentes : d'une part, la question de l'étendue des inégalités ; d'autre part, la question de la justification des inégalités.

On pourrait avoir l'impression que la première question renvoie à ce que l'on entend habituellement par la question de l'égalité, et que la deuxième question renvoie à la question du mérite. On pourrait alors avoir une société fortement égalitaire et méritocratique, fortement égalitaire et non-méritocratique, fortement inégalitaire et méritocratique, ou bien fortement inégalitaire et non-méritocratique, en fonction de la variation de l'écart entre riches et pauvres,

²⁹⁴ Dans les faits, comme l'ont bien démontré un certain nombre d'enquêtes en économie, le taux de mobilité sociale est relativement faible aux États-Unis ; il est même, en moyenne, plus faible qu'en Europe. Je parle bien ici du stéréotype associé aux États-Unis comme pays de liberté et de réalisation personnelle. Shai DAVIDAI et Thomas GILOVICH, « Building a More Mobile America—One Income Quintile at a Time », *Perspectives on Psychological Science*, 2015, vol. 10, n° 1, p. 60-71.

et en fonction des facteurs qui expliquent cette différence de revenus. On pourrait alors penser qu'il n'y a fondamentalement aucun rapport entre la question du mérite et la question de l'égalité, et que les analyses politiques classiques, qui ont cherché à lier les deux questions, sont erronées ; une analyse philosophique correcte ferait alors voir qu'il n'existe aucun lien nécessaire entre les deux questions. Le mérite ne serait pas plus l'ennemi de l'égalité qu'il n'est en contradiction avec la sécurité ou le bien-être. Il s'agirait simplement de questions à traiter séparément.

Même s'il existe sans doute une part de vérité dans l'analyse qui précède, je défendrai ici l'idée qu'il est illusoire de penser que la question du mérite est sans rapport avec la question de l'égalité. En réalité, la satisfaction du mérite est indispensable pour promouvoir l'égalité sociale. Les deux notions ne sont pas opposées, contrairement à ce que l'on a longtemps prétendu ; elles ne sont pas non plus indépendantes ; les deux idées sont alliées, et le mérite est le fondement d'une possible réalisation de l'égalité sociale au sein de notre société. La représentation la plus flagrante d'une société inégalitaire n'est pas l'image d'une société méritocratique, mais c'est au contraire l'image d'une société de castes, de privilèges, de distinctions illégitimes. Une société où règne le népotisme et la corruption représente le cœur de ce que l'on entend par une société inégalitaire. S'il existe des inégalités au sein de la société, le fait que ces inégalités soient en partie justifiées par le mérite respectif des individus est la seule manière de préserver la dignité de chaque citoyen.

Le dernier paragraphe a été écrit dans le but de provoquer le lecteur. En réalité, mérite et égalité peuvent être ou bien opposés, ou bien indépendants, ou bien alliés, selon la définition que l'on prend des deux notions. Je tenterai de montrer ici que, si l'on considère la définition du mérite et de l'égalité qui ont le plus de poids d'un point de vue normatif, alors les deux notions sont fondamentalement inséparables, la réalisation du mérite au sein de la société étant la condition de possibilité de l'égalité sociale. Cependant, pour voir en quoi différentes manières de concevoir les deux notions peuvent changer la vision que l'on a des deux notions, il faut procéder à une série de distinctions entre différentes acceptations de ces deux idées.

Pour approfondir la relation entre mérite et égalité, il est important de faire la distinction entre deux sens de l'idée d'égalité : l'égalité distributive et l'égalité sociale. L'égalité distributive correspond à la distribution égale d'un bien particulier. Par exemple, l'égalité distributive dans le domaine de l'argent pourrait correspondre à un revenu égal pour tous, ou à une situation avec un écart de revenu très faible. L'égalité sociale correspond à ce qui a été également appelé égalité relationnelle, ou égalité complexe par certains auteurs. Alors que l'égalité distributive caractérise essentiellement le degré ou l'étendue de l'écart de revenu entre

différents membres d'une société, l'égalité sociale renvoie à l'état des relations entre différents individus au sein de la société, et décrit la mesure dans laquelle ces relations sont caractérisées par des relations de domination, de hiérarchie, ou d'égalité.

De nombreux auteurs ont conceptualisé l'idée d'égalité sociale de différentes manières ; j'indiquerai ici simplement certains traits qui reviennent souvent dans la description de l'égalité sociale. L'égalité sociale correspond à l'idée d'une société où auraient été éliminées les différences de statut ou de privilège, et où règnerait une égalité de respect au sein des citoyens. On a ainsi pu caractériser l'égalité sociale comme correspondant à une absence de domination au sein d'un groupe social²⁹⁵.

A partir de cette distinction, il est possible de voir deux types de relation possible entre mérite et égalité. Si l'on prend l'idée d'égalité comme égalité distributive, alors les idées de mérite et d'égalité sont largement indépendantes, comme nous l'avons vu. Alors que l'égalité distributive caractérise essentiellement le degré ou l'étendue de l'écart de revenu entre différents membres d'une société, l'idée de mérite renvoie aux raisons expliquant ou justifiant l'existence de cet écart. Dans ce cadre, comme indiqué plus haut, il est possible d'imaginer toutes les variations entre ces deux idées : un écart faible ou élevé, qui s'explique pour des raisons de mérite ou non.

A l'inverse, si l'on comprend l'idée d'égalité comme égalité sociale, alors la satisfaction du mérite au sein d'une société est la condition de réalisation de l'égalité. Tant qu'il existe un écart de revenu au sein d'un groupe, le fait que cet écart s'explique par un mérite différent est la seule raison compatible avec l'égalité de toutes les parties prenantes. Si cet écart s'explique par des raisons de népotisme, de privilège, ou de corruption, alors il semble difficile de concevoir les différents membres de la société comme égaux. Des règles différentes s'appliquent à différents membres, traduisant l'idée que certains disposent d'une plus grande valeur que d'autres. L'idée d'égalité sociale présuppose que tous les membres d'une société soient soumis au même statut, et la récompense en fonction du mérite est le seul critère permettant à tous les individus de se concevoir comme soumis à un espace commun. Si les inégalités de récompense traduisent des différences de contribution, alors il existe une transparence dans la soumission à des règles communes.

Ces propositions pourront sembler familières au lecteur. En effet, il s'agit d'une simple reprise d'une idée fondamentale à l'œuvre depuis les révolutions française et américaine, et qui est restée au cœur de nos idéaux politiques depuis lors. Cette idée, c'est que l'égalité est avant tout une lutte contre les privilèges et contre la hiérarchie sociale, comprise comme légitimation

²⁹⁵ M. WALZER, *Spheres Of Justice*, *op. cit.* ; D. MILLER, *Principles of Social Justice*, *op. cit.* ; E. S. ANDERSON, « What Is the Point of Equality? », *op. cit.* ; J. WOLFF, « Fairness, Respect, and the Egalitarian Ethos », *op. cit.*

d'une différence intrinsèque entre les différents groupes sociaux. Parce que le mérite est avant tout basé sur la contribution, et non sur la naissance, il permet la création d'un monde commun où tous disposent des mêmes droits. Cette idée est donc bien vieille. Mais il peut être important de la rappeler, dans un monde où l'on se complait généralement à ressasser sans cesse l'opposition irréductible entre mérite et égalité.

Bien sûr, il faut rester prudent dans cette défense d'une harmonie entre les idéaux d'égalité et de mérite. Et ce, pour trois raisons. Tout d'abord, l'idée de mérite laisse ouverte la possibilité d'inégalités justes. En ce sens, le mérite est bien sûr opposé à l'égalité, même s'il ne dit rien sur l'étendue des inégalités qui peuvent être ainsi être justifiées. Ensuite, il est évident que l'idée de mérite peut être détournée, et que certaines personnes ou certains partis politiques peuvent se servir de cette idée pour justifier des inégalités héritées qui n'ont rien à voir avec le travail, la contribution ou les efforts. Enfin, troisièmement, comme Michael Young y insistait quand il forgea le terme de méritocratie, il est possible d'attacher trop d'importance à des différences de mérite, et de tout ramener à l'auge du travail productif. Bien sûr, le travail a son importance. Mais le mérite n'est qu'une partie de la justice, qui n'est elle-même qu'une partie de la moralité, qui n'est elle-même qu'une partie de la vie bonne. Attacher trop d'importance au mérite risque de nous faire perdre de vue d'autres facteurs, qui eux, valorisent l'égalité pour elle-même, comme l'attachement à la citoyenneté.

Dans le monde d'aujourd'hui, il semble assez clair que le mérite et l'égalité vont dans le même sens, celui d'une remise en cause des inégalités radicales de richesse. Dans un pays où l'héritage joue un tel rôle, il semble bien que la sacralisation du droit de propriété représente une bien plus grande menace pour l'égalité que le mérite.

En somme, étant donné que nous disposons d'une pluralité de justifications du mérite, qu'il s'agit d'une conception extrêmement populaire, qui n'exige pas de conception trop exigeante de la responsabilité pour être plausible, et qui n'entre pas en fort conflit avec nos autres principes moraux, il semblerait difficile de voir pourquoi nous devrions renoncer à cette idée. Parfois, il est préférable de conserver nos idées courantes. Dans le cas du mérite, le sens commun semble nous renvoyer un idéal qui vaut encore la peine d'être poursuivi.

Appendice général

Méthodes expérimentales et problèmes de généralisation

Trois chapitres de cette thèse offrent la présentation de résultats empiriques portant sur les perceptions ordinaires du mérite. Comme indiqué dans l'introduction, de telles études sont devenues relativement courantes au cours des 20 dernières années, avec l'expansion considérable du champ de la philosophie expérimentale. Je souhaiterais indiquer ici quelques considérations méthodologiques qui me semblent indispensables pour appréhender ces études. Cette thèse ne s'adressant évidemment pas uniquement, ou même essentiellement, à des chercheurs familiers avec les méthodes quantitatives utilisées en psychologie sociale, il semble important d'étudier plus longuement quatre problèmes méthodologiques essentiels. Tout d'abord, je ferai de brefs rappels sur l'usage de statistiques inférentielles, et sur la possibilité de faire des inférences à partir d'un échantillon restreint étudié. Dans un deuxième temps, j'indiquerai brièvement comment lire les graphiques utilisés dans cette thèse. Dans un troisième temps, j'étudierai le problème posé par le recrutement des échantillons classiques en psychologie sociale, à savoir des populations d'étudiants ou de personnes recrutées sur internet. Enfin, dans un quatrième temps, j'effectuerai un très bref rappel des problèmes contemporains posés par la crise de la répliquabilité en sciences sociales, et les moyens que j'ai utilisés pour contrer ces difficultés.

De l'usage des statistiques inférentielles dans les disciplines quantitatives

Je vais ici faire une brève présentation des méthodes statistiques employées dans cette thèse, à destination du lecteur qui n'aurait pas eu de formation en méthodes quantitatives. Même si l'ensemble des chapitres quantitatifs de cette thèse expliqueront naturellement ce que signifient

les résultats, il est préférable que le lecteur soit un minimum familier avec l'usage des statistiques utilisées.

Une distinction fondamentale est la différence entre statistiques descriptives et statistiques inférentielles. Pour comprendre cette distinction, il faut bien garder à l'esprit l'écart entre l'échantillon recruté et la population à laquelle on souhaite généraliser. Si l'on recrute 30 citoyens américains pris au hasard, alors nous disposons d'un échantillon de trente personnes, et la population à laquelle nous désirons généraliser est la population constituée par l'ensemble des citoyens américains.

Dans ce cadre, les statistiques descriptives correspondent à la description des caractéristiques de l'échantillon, sans que l'on se pose le moindre du monde la question de savoir si on peut généraliser ces données à l'ensemble de la population. A l'inverse, les statistiques inférentielles renvoient au fait d'effectuer une généralisation à l'ensemble de la population cible. Par exemple, si l'on recrute trente participants et que l'on s'aperçoit qu'ils ont en moyenne 37 ans, alors cet âge moyen constitue une statistique descriptive : il s'agit simplement d'une description de l'échantillon, sans que l'on puisse en inférer que la population américaine ait en moyenne 37 ans. Si l'on veut faire une généralisation, alors il faudra accepter l'incertitude dans nos estimations. En effet, il se peut que nous ayons par hasard recruté trente personnes qui sont légèrement plus jeunes ou légèrement plus âgées que l'âge moyen de la population américaine. Dans ce cadre, on peut faire des estimations révélant l'incertitude de nos données : par exemple, avec l'aide de certaines hypothèses auxiliaires, on peut construire un intervalle de confiance, qui nous indique les valeurs « probables » de l'âge moyen au sein de la population américaine²⁹⁶.

Avec cette distinction entre statistiques descriptives et inférentielles en place, voyons un exemple de statistiques qui seront utilisées dans la suite de la thèse. Imaginons que le lecteur lise une expérience décrivant la différence entre deux conditions expérimentales. Dans une condition, les participants de l'expérience ont assisté à une punition où le punisseur avait clairement fait savoir pourquoi il infligeait une punition. Dans l'autre condition, le punisseur avait agi de façon anonyme, si bien que la personne punie n'avait pas compris pourquoi elle l'avait été. Les participants devaient ensuite noter de 1 à 7 la justice de la punition, 1 indiquant une punition très injuste, 7 une punition très juste. Imaginons que le lecteur rencontre alors les valeurs suivantes comme résumé des réponses des participants : $\Delta M=0,73$, 95% CI [0,32, 1,15],

²⁹⁶ L'interprétation d'un intervalle de confiance est une question complexe et débattue.

$t(191,68)=3,49$, $p=,001$. A première vue, ces indications peuvent sembler incompréhensibles. Cependant, une considération de chacun des termes séparément devrait éclairer le lecteur.

« $\Delta M=0,73$ » est une statistique descriptive. Il s'agit de la différence moyenne entre les deux conditions. En l'occurrence, la moyenne d'une des deux conditions était de 0,73 points au-dessus de la moyenne de l'autre condition. Le fait que cette différence moyenne a été obtenue ne dit cependant pas qu'une telle différence se retrouverait au sein de la population générale. Pour pouvoir généraliser à la population d'ensemble, il faut passer à des statistiques inférentielles.

« 95% CI [0,32, 1,15] » correspond à l'intervalle de confiance à 95% de la différence de moyenne entre les deux conditions. Un intervalle de confiance à 95% correspond au fait que, si l'on tirait des échantillons de la population cible une infinité de fois, alors la véritable différence moyenne entre les deux conditions se trouverait au sein de l'intervalle de confiance 95% du temps. Si l'on ne dispose d'aucune autre information sur la différence réelle entre les deux conditions, alors il existe une probabilité de 95% que la véritable différence se trouve au sein de l'intervalle de confiance. De plus, toujours en l'absence d'autre information, les valeurs réelles les plus probables sont celles tournant autour de la moyenne trouvée au sein de l'échantillon (dans notre cas, 0,73) ; les valeurs aux extrémités de l'intervalle de confiance (0,32 et 1,15) sont en général 7 fois moins probables que les valeurs proches de 0,73²⁹⁷.

La troisième information importante dans cette description est la valeur-p, indiquée ici par « $p=,001$ ». La valeur-p est construite à partir de l'opposition entre deux hypothèses, l'hypothèse nulle et l'hypothèse alternative. L'hypothèse nulle est l'hypothèse que l'expérimentateur cherche à falsifier ; la plupart du temps, il s'agit de l'hypothèse qu'il n'existe aucune différence entre la condition contrôle et la condition expérimentale. L'hypothèse alternative correspond à l'ensemble des autres possibilités qui sont incompatibles avec l'hypothèse nulle (la plupart du temps, cela correspond donc à l'hypothèse qu'il existe une différence entre la condition contrôle et la condition expérimentale. Selon la présentation proposée par le statisticien et psychologue Eric-Jan Wagenmakers, on peut concevoir l'hypothèse nulle (l'absence de différence) comme l'hypothèse par défaut : on peut toujours supposer qu'une manipulation expérimentale n'a aucun impact sur le comportement des individus²⁹⁸. Le chercheur doit alors démontrer que ce défaut ne s'applique pas, et que la

²⁹⁷ Geoff CUMMING, *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis.*, New York, Routledge, 2011.

²⁹⁸ Eric-Jan WAGENMAKERS et Quentin GRONAU, « Redefine Statistical Significance Part X: Why the Point-Null Will Never Die ».

manipulation a bien une influence. Dans ce cadre, la valeur-p renvoie à la probabilité d'obtenir des résultats aussi extrêmes que ceux que l'on a obtenus si jamais l'hypothèse nulle était vraie. On imagine donc qu'il n'y a aucune différence entre les deux conditions, et l'on se demande alors dans quelle mesure des résultats aussi extrêmes que ceux obtenus arriveraient fréquemment dans cette condition. Si des résultats aussi extrêmes n'arriveraient que très rarement (moins d'une sur vingt, selon le seuil habituellement retenu), alors on décide de rejeter l'hypothèse nulle, et de considérer qu'il existe effectivement une différence moyenne entre les deux conditions. Dans le cas qui nous occupe, « $p=,001$ » correspond au fait que des différences aussi extrêmes ne seraient observées qu'une fois sur mille s'il n'y avait pas de différence entre les deux conditions. Bien sûr, il est toujours possible que l'hypothèse nulle soit vraie, qu'il n'y ait aucune différence, mais qu'un événement très rare se soit produit²⁹⁹.

Enfin, « $t(191,68)=3,49$ » indique que nous nous sommes servis d'un t-test pour effectuer nos inférences statistiques. Le t-test renvoie à la méthode utilisée pour construire notre intervalle de confiance et notre valeur-p. Même s'il s'agit d'un facteur important pour évaluer dans quelle mesure les méthodes utilisées sont correctes, le lecteur néophyte en statistiques devrait probablement négliger cette information, et se concentrer sur la différence de moyenne, l'intervalle de confiance, et la valeur-p.

Une brève indication sur la lecture des graphiques

Les graphiques dont je me servirai au cours des pages suivantes possèdent certaines spécificités par rapport aux graphiques habituellement utilisés, en tant qu'ils montrent légèrement plus d'informations. Je me servirai du graphique suivant, que le lecteur pourra retrouver dans le chapitre 8, pour indiquer en général les informations déployées.

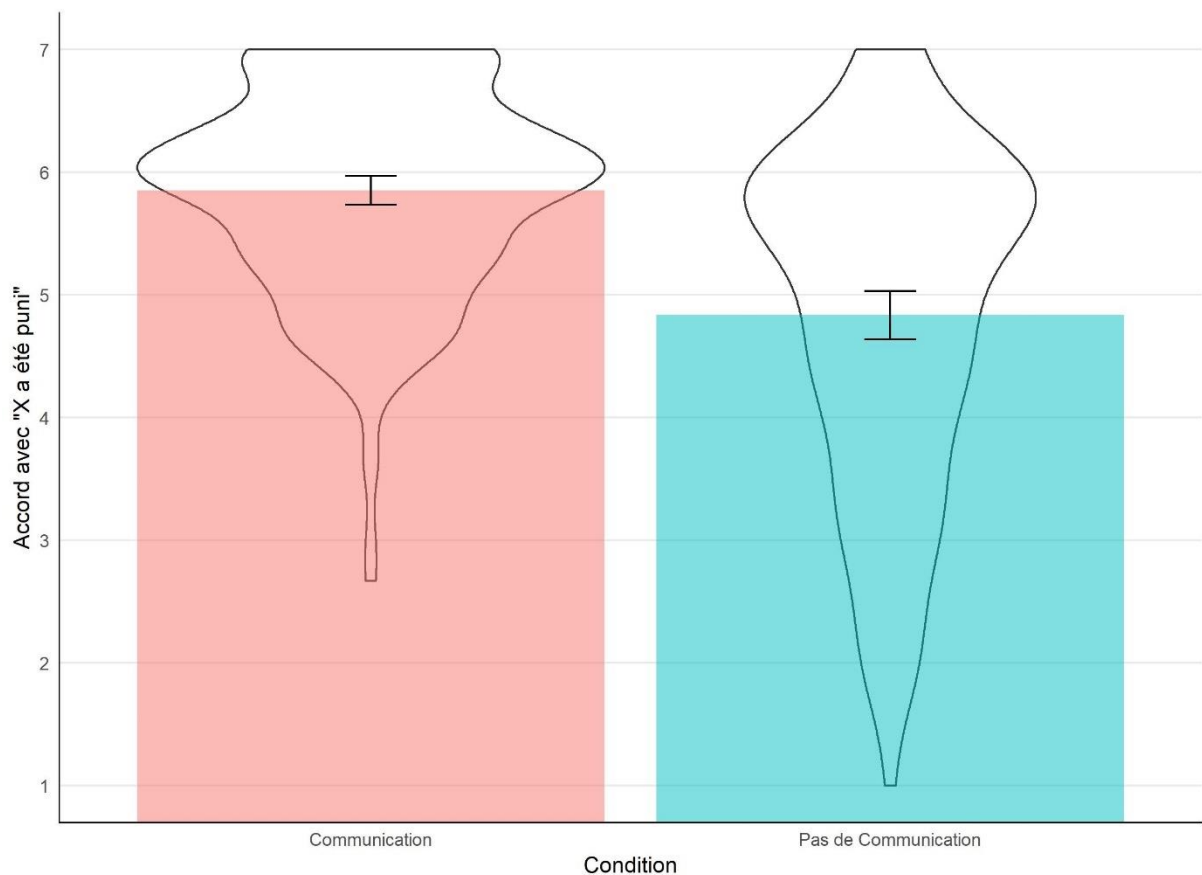
Tout d'abord, le graphique suivant peut être lu de la façon la plus simple possible en le considérant comme un exemple d'un diagramme à barres. Dans ce cas, chacune des barres colorées représente une condition expérimentale, et il est possible de lire la moyenne dans chacune des conditions en regardant jusqu'où monte chacune des barres. Dans le cas de ce

²⁹⁹ L'usage des valeurs-p est extrêmement controversée, et a fait l'objet de nombreuses critiques tout au long de l'histoire des disciplines quantitatives. Il s'agit cependant d'une méthode utile pour limiter le risque d'erreur dans nos inférences. Cf. R. S. NICKERSON, « Null hypothesis significance testing: a review of an old and continuing controversy », *Psychological Methods*, 2000, vol. 5, n° 2, p. 241-301.

graphique, on voit que la condition *Communication* (en rose) a une moyenne un peu en dessous de 6, et la moyenne *Pas de communication* a une moyenne un peu en dessous de 5.

Ensuite, il est possible de complexifier la lecture en faisant attention aux flèches noires qui se situent autour du sommet de chacune des barres colorées. Ces flèches noires indiquent l'intervalle de confiance à 95% de chacune des conditions. Dans le cas de la condition *Communication* (en rose), l'intervalle de confiance est à peu près compris entre 5,8 et 6.

Enfin, le lecteur pourra remarquer le graphique de densité, en noir, qui monte jusqu'à 7 dans les deux conditions, descend un peu en dessous de 3 dans le cas de la condition *Communication*, et descend à 1 dans le cas de la condition *Pas de communication*. Ce graphique indique la densité de réponses. La largeur de chaque graphique de densité est proportionnelle au nombre de participants ayant donné chacune de ces réponses. Dans le cas de la condition *Communication*, on voit que le graphique de densité est le plus large un peu au-dessus de 6, ce qui indique que ce type de réponse est celui qui a été le plus populaire au sein des participants. A l'inverse, on voit que très peu de participants ont donné des réponses en-dessous de 4, et personne en dessous de 2,7. Donner la densité des réponses permet d'obtenir davantage d'informations sur le comportement, au-delà de la simple moyenne des réponses.



Les étudiants et internet comme sources d'information : problèmes et promesses

Les expériences suivantes ont été effectuées au sein de deux publics différents : des étudiants de l'université de Pennsylvanie, aux États-Unis, et des personnes recrutées au sein de la plateforme internet *Amazon Mechanical Turk*. *Amazon Mechanical Turk* (*Mturk* à partir de maintenant) est un mode de recrutement de participants de plus en plus utilisé par les chercheurs en sciences sociales, en tant qu'il s'agit d'un moyen peu coûteux de recruter un très grand nombre de participants. Aussi bien dans le cas des étudiants que dans le cas de *Mturk*, il ne s'agit évidemment pas d'un échantillon probabiliste, l'étalon-or du recrutement de participants³⁰⁰. Le lecteur est en droit de se demander dans quelle mesure des échantillons aussi spécifiques peuvent nous apporter des connaissances sur la population américaine en général ; de la même manière, même en faisant l'hypothèse que les résultats pourraient être généralisés à l'ensemble de la population américaine, on peut se demander dans quelle mesure ils s'appliqueraient également aux populations européennes, voire à d'autres régions géographiques. Se poser cette question est d'autant plus importante que les participants à *Mturk* ont légèrement plus de chances d'être des hommes, d'être jeunes, éduqués, et de gauche, que des participants tirés au hasard au sein de la population américaine³⁰¹.

En ce qui concerne la généralisation à la population américaine en son ensemble, les données disponibles sont étonnamment positives. Plusieurs chercheurs en sciences politiques ont récemment cherché à comparer les performances du recrutement de sujets via *Amazon Mechanical Turk* au recrutement de sujets par des moyens probabilistes. Les conclusions générales sont que *Mturk* fournit des échantillons de très bonne qualité, relativement proches de ce qu'aurait donné un moyen de recrutement probabiliste. Il faut, cependant, rester prudent sur ce point, notamment à partir de la distinction entre la taille d'effet d'une manipulation expérimentale, et les résultats absolus obtenus au sein de chaque condition.

Pour détailler la différence entre chiffres absolus au sein de chaque condition et taille d'effet entre deux conditions, prenons l'exemple d'une expérience qui tenterait d'évaluer l'impact d'un

³⁰⁰ Un échantillon recruté de façon probabiliste est un échantillon où chaque personne au sein d'une population avait *a priori* une même probabilité d'être recrutée. Il s'agit d'un mode de recrutement extrêmement compliqué à mettre en place, et en conséquence très coûteux.

³⁰¹ A l'inverse, la diversité géographique de *Mturk*, entre les différentes régions des États-Unis, est remarquable, et concorde très fortement avec les données géographiques objectives. Sharad GOEL, Adam OBENG et David ROTHSCHILD, « Non-Representative Surveys: Fast, Cheap, and Mostly Accurate », *manuscrit non publié*.

texte insistant sur les coûts de l'immigration sur le soutien à la politique migratoire de Donald Trump³⁰². La condition expérimentale reçoit ce texte anti-immigration, puis une question demandant s'ils soutiennent ou non la politique du président américain sur le sujet. La condition contrôle contient un texte n'ayant rien à voir avec l'immigration, puis reçoit la même question que la condition expérimentale. Imaginons maintenant que cette expérience ait été faite aussi bien sur *Amazon Mechanical Turk* qu'auprès d'un échantillon représentatif. On peut alors se poser deux questions à ce sujet. D'une part, est-ce que les chiffres absolus d'accord avec le président américain au sein de chaque condition sont concordants entre l'échantillon de *Mturk* et l'échantillon représentatif ? Ensuite, la manipulation expérimentale d'insistance sur les coûts de l'immigration a-t-elle eu le même effet sur l'échantillon de *Mturk* et sur l'échantillon représentatif ? En ce qui concerne les chiffres absolus au sein de chaque condition, étant donné que l'échantillon de *Mturk* est connu comme étant plus à gauche que la population américaine dans son ensemble, on peut supposer qu'il y aura davantage d'opposition que dans le cas de la population américaine. Étant donné la relative impopularité du président américain à l'heure actuelle, on peut imaginer que 60% des américains se déclareraient opposés aux politiques migratoires du président dans l'échantillon représentatif, et 70% sur *Mturk*. Il s'agirait alors d'une différence notable. Cela ne veut cependant pas dire que les deux échantillons réagiraient de manière différente à la manipulation expérimentale. On peut très bien imaginer que, dans les deux cas, insister sur les coûts de l'immigration conduirait une partie des gens à se déclarer favorables aux politiques de Donald Trump, et que ce revirement connaîtrait une force similaire aussi bien sur *Mturk* que sur un échantillon probabiliste.

De fait, les études comparant la taille des manipulations expérimentales sur des échantillons probabilistes et sur *Mturk* concluent souvent que les résultats sont similaires dans les deux cas. Le chercheur en sciences politiques Alexander Coppock en montre un effet frappant : sur 10 expériences faites aussi bien sur *Mturk* que sur des échantillons probabilistes, la corrélation entre les tailles d'effets obtenus sur *Mturk* et sur les échantillons probabilistes a été de 0,96, soit presque un accord parfait³⁰³. Une autre comparaison, basée sur un échantillon plus large de 20 expériences, concluait également à un accord très fort entre les effets trouvés sur *Mturk* et sur

³⁰² Je prends cet exemple puisqu'il s'agit d'une des expériences utilisées par Kevin J. MULLINIX, Thomas J. LEEPER, James N. DRUCKMAN et Jeremy FREESE, « The Generalizability of Survey Experiments », *Journal of Experimental Political Science*, 2015, vol. 2, n° 2, p. 109-138. L'exemple dont je me sers est cependant simplifié par rapport à la véritable expérience.

³⁰³ Alexander COPPOCK, « Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach », *Political Science Research and Methods*, 2018, p. 1-16.

un échantillon représentatif³⁰⁴. Même s'il faut rester prudent sur ce point, étant donné que les études disponibles restent limitées, il semble bien qu'un effet obtenu sur *Mturk* se généralise très bien sur la population américaine dans son ensemble.

La question est plus compliquée en ce qui concerne l'accord entre les pourcentages absolus obtenus sur *Mturk* et sur des échantillons représentatifs. Sharad Goel et ses collègues ont conduit une étude récente qui montre qu'il existe un fort accord entre *Mturk* et les échantillons probabilistes sur ce point également³⁰⁵. Goel et ses collègues ont interrogé les participants sur 49 questions portant sur des sujets politiques et sociaux sensibles, comme par exemple la croyance en Dieu ou les attitudes relatives à l'homosexualité. Étant donné le biais à gauche de *Amazon Mechanical Turk*, il s'agit des questions pour lesquelles on pourrait s'attendre au plus fort écart entre les deux types de méthode. Dans ce cas, l'écart médian entre les pourcentages obtenus par les deux méthodes a été de 9,1 points de pourcentage. Les auteurs interprètent cet écart comme un écart assez faible ; il est cependant clair que la différence n'est pas nulle. *Mturk* donne donc une image assez fidèle de la société américaine, même si cette image n'est pas sans erreur. Dans le cas des expériences qui nous occuperont dans cette thèse, il s'agit de questions qui sont beaucoup moins politisées et controversées que les questions liées à l'homosexualité ou au rôle de l'État dans l'économie. Je pense en conséquence que les résultats obtenus sont beaucoup plus proches de la vérité au sein de la population américaine en son ensemble que dans le cas des questions sensibles étudiées par Goel et ses collègues. Toute inférence doit cependant rester extrêmement prudente, et le lecteur pourra garder en tête que les résultats pourraient varier de plusieurs points de pourcentage avec un autre type de recrutement de participants.

L'extrapolation de la population américaine aux populations d'autres pays est une question encore plus complexe. Je crois personnellement que toutes les expériences effectuées dans le cadre de cette thèse donneraient des résultats similaires en France ou dans d'autres pays européens ; cependant, ne disposant pas de données sur le sujet, je ne demanderai évidemment pas au lecteur de me croire sur parole. Il faut cependant remarquer que les effets expérimentaux sont souvent similaires des deux côtés de l'Atlantique. Nous disposons maintenant de bonnes données sur le sujet, grâce aux récentes collaborations internationales dites *Many Labs*, qui ont vu la réplique de 41 expériences dans plus de 80 laboratoires différents, situés en Europe, en

³⁰⁴ K. J. MULLINIX, T. J. LEEPER, J. N. DRUCKMAN et J. FREESE, « The Generalizability of Survey Experiments », *op. cit.*

³⁰⁵ S. GOEL, A. OBENG et D. ROTHSCHILD, « Non-Representative Surveys: Fast, Cheap, and Mostly Accurate », *op. cit.*

Amérique du Nord, et dans de nombreux pays d'Amérique du Sud et d'Asie. Les 41 expériences ont donné des résultats très proches dans tous les pays³⁰⁶. Cependant, il ne s'agit pas de nier les différences culturelles, ou le fait que celles-ci puissent avoir un impact sur les réponses des participants ; la généralisation des expériences faites ici à d'autres populations est une simple hypothèse.

La crise de la répliquabilité en sciences sociales

Les dix dernières années ont constitué un changement historique radical dans les pratiques de recherche effectuées en psychologie et dans une partie des sciences sociales. Ce mouvement, à l'œuvre depuis 2011, a été décrit de diverses manières : parfois de façon négative, sous le titre de « crise de la répliquabilité », parfois de façon positive, avec l'idée que ces années ont constitué « la renaissance de la psychologie » comme discipline scientifique³⁰⁷. Une description complète de ce changement historique est évidemment hors de question ici ; je me permets de renvoyer le lecteur à l'article d'introduction générale que j'ai publié à la *Vie des idées*³⁰⁸.

Depuis 2011, un mouvement général de répliquabilité des expériences passées a eu lieu en psychologie sociale et cognitive. Une tentative de répliquabilité consiste à refaire une enquête ou une expérience sur un nouvel échantillon, à partir des mêmes méthodes utilisées dans l'expérience ou l'enquête originale. Si les résultats sont concordants, alors on peut parler d'une répliquabilité réussie ; si les résultats divergent fortement, on peut dire qu'il y a un échec de la répliquabilité. Ce grand mouvement de tentative de répliquabilité a abouti à la découverte que de très nombreuses expériences en psychologie sociale et cognitive ne sont pas répliquables. Dans de nombreux cas, les chercheurs tentant de répliquer les expériences de leurs collègues aboutissent à des résultats radicalement différents. Ainsi, le *Reproducibility Project : Psychology*, un consortium de plusieurs centaines de chercheurs, a abouti à la répliquabilité réussie de 39 des 100

³⁰⁶ Richard A. KLEIN, Kate A. RATLIFF, Michelangelo VIANELLO, Reginald B. ADAMS JR., Štěpán BAHNÍK, Michael J. BERNSTEIN, ... et Brian A. NOSEK, « Investigating variation in replicability: A “many labs” replication project », *Social Psychology*, 2014, vol. 45, n° 3, p. 142-152 ; Richard Anthony KLEIN, Michelangelo VIANELLO, Fred HASSELMAN, Byron Gregory ADAMS, Reginald B. ADAMS, Sinan ALPER, ... et Brian A. NOSEK, « Many Labs 2: Investigating Variation in Replicability Across Sample and Setting ».

³⁰⁷ Leif D. NELSON, Joseph SIMMONS et Uri SIMONSOHN, « Psychology's Renaissance », *Annual Review of Psychology*, 2018, vol. 69, n° 1, p. null.

³⁰⁸ Aurélien ALLARD, « La crise de la répliquabilité », *La Vie des idées*, 2018. Les paragraphes suivants reprennent assez largement les conclusions de cet article. Pour une perspective plus complète sur la question, voir notamment le livre de Chris CHAMBERS, *The Seven Deadly Sins of Psychology: A Manifesto for Reforming the Culture of Scientific Practice*, Princeton University Press, 2017.

expériences sélectionnées à partir d'articles publiés dans des revues prestigieuses de psychologie sociale et cognitive. Un tel échec a lancé un signal d'alarme, qui a conduit à un renouvellement radical des méthodes dans le domaine.

Avant de détailler davantage les causes de ce manque de répliquabilité en psychologie sociale, il faut souligner qu'il s'agit d'un problème généralisé à l'ensemble des sciences sociales, biologiques, et bio-médicales, et non pas un problème propre à la psychologie. Par exemple, un projet de répliquabilité d'un ensemble d'expériences en biologie du cancer est en cours, et les résultats ont été pour l'instant extrêmement mitigés³⁰⁹. Les études publiées en expérimentation animale sont souvent conduites avec de tous petits échantillons, et ne respectent souvent pas les règles minimales de rigueur dans le processus d'attribution aléatoire des sujets aux différentes conditions expérimentales³¹⁰. Bien sûr, il existe de très grandes variations dans la qualité des différents champs scientifiques. Au sein des différentes disciplines médicales, les études effectuées sur les humains sont souvent beaucoup plus contrôlées, et sont beaucoup plus rigoureuses, que les études effectuées sur les animaux.

On peut distinguer 4 facteurs principaux à l'origine du manque de fiabilité des résultats publiés dans les revues de sciences sociales et de recherche bio-médicale : le fait que la culture scientifique accorde une importance exclusive aux résultats novateurs et « positifs », l'excessive flexibilité dans l'analyse statistique des résultats, l'absence de répliquabilité des travaux antérieurs par des laboratoires autres que ceux ayant publié les études princeps, et la faiblesse des échantillons.

La culture scientifique dans les sciences sociales et bio-médicales accorde une forte valeur aux résultats novateurs et positifs. Les revues scientifiques préféreront publier des articles portant sur un nouveau phénomène, et tout particulièrement si ceux-ci montrent que ce phénomène est porteur d'applications pratiques décisives. La principale conséquence de cet état d'esprit est que les chercheurs sont incités à ne pas publier de résultats « négatifs », c'est-à-dire des résultats n'indiquant aucune différence significative entre divers traitements. Par exemple, dans le domaine médical, on préférera publier un article portant sur un traitement prometteur du cancer qu'un article décrivant comment un nouveau traitement échoue à le guérir. Cela a pour conséquence que seule une partie des expériences sont publiées, ce qui aboutit à ce qu'on appelle un « biais de publication ».

³⁰⁹ Jocelyn KAISER, « Plan to replicate 50 high-impact cancer papers shrinks to just 18 », *Science / AAAS*, 2018, 31/07/2018 p.

³¹⁰ David W. HOWELLS, Emily S. SENA et Malcolm R. MACLEOD, « Bringing rigour to translational medicine », *Nature Reviews Neurology*, 2014, vol. 10, n° 1, p. 37-43.

Un tel biais peut conduire à la publication exclusive d'articles démontrant une différence, même là où il n'en existe aucune. Imaginons que 20 laboratoires médicaux cherchent à déterminer si une molécule particulière peut guérir un cancer. Il arrivera nécessairement que certains laboratoires découvriront, par pur hasard, une relation positive entre ce traitement et la cure du cancer (par exemple, parce que les personnes traitées avaient au préalable un cancer moins agressif que les personnes dans le groupe témoin). D'autres découvriront peut-être une relation négative, donnant l'impression que le traitement aggrave la situation. En fin de compte, seul le laboratoire qui aura trouvé un effet fort du traitement sur la guérison du cancer pourra publier son article, parce que seul celui-ci semblera suffisamment intéressant. Les 19 autres laboratoires, à l'inverse, relégueront probablement leur étude au fond d'un tiroir, parce qu'ils savent que, si cette étude ne promet pas de guérir le cancer, elle n'est pas publiable.

La conséquence de telles pratiques est que la recherche publiée dans les revues scientifiques exagère fondamentalement l'efficacité des traitements médicaux, des interventions pédagogiques ou éducatives, ou, dans le domaine psychologique, l'influence de manipulations subtiles sur le comportement humain. La littérature scientifique se retrouve alors parsemée de résultats faux, ou, au minimum, très largement exagérés.

Ce biais de publication est d'autant plus problématique qu'il tend à inciter les chercheurs à « arranger » leurs résultats. Pour maximiser leurs chances de publication, ils peuvent avoir recours à des pratiques de recherche discutables, diminuant la rigueur de la recherche, mais maximisant les chances d'obtenir un résultat positif. Une première pratique discutable consiste à recruter des participants progressivement, en plusieurs étapes, dans le but d'arrêter l'expérience uniquement quand les résultats voulus ont été obtenus. Une deuxième pratique courante consiste à tester plusieurs conditions en même temps, dans le but de comparer ces multiples résultats possibles, avant de se concentrer sur la seule expérience qui a « marché ». Un troisième type de pratiques discutables consiste à effectuer des analyses sur une seule partie du groupe testé. Par exemple, dans la recherche bio-médicale, si des chercheurs obtiennent des résultats qui indiquent qu'un nouveau médicament n'est pas très efficace, ils peuvent chercher à subdiviser leurs analyses pour voir s'il n'existe pas un effet positif sur les femmes, les hommes, les personnes jeunes, âgées, etc. À force de diviser la population en multiples sous-groupes, il y aura une forte chance que le médicament se trouve, par pur hasard, profiter à un groupe particulier, même s'il n'est en réalité bénéfique pour personne.

Ce manque de fiabilité de la recherche initiale ne serait pas en soi un problème gigantesque si la recherche scientifique se corrigeait au fil du temps. En sciences, une telle correction peut prendre la forme de tentatives de réplification d'expériences précédentes. Une telle pratique est

extrêmement courante en physique, par exemple. Répliquer une expérience antérieure est devenue relativement courant en psychologie depuis le début de la crise de la répliquabilité en 2011. Mais cette pratique était, jusque-là, quasiment inexistante : il était extrêmement difficile de parvenir à faire publier une réplification, puisque de tels articles n'étaient pas considérés comme suffisamment novateurs. En conséquence, des études parfaitement erronées ont pu continuer d'être citées pendant des décennies sans qu'aucune correction y soit apportée.

Dernier facteur majeur du manque de fiabilité, et non des moindres : la faiblesse des échantillons. Les expériences sont en effet conduites sur un nombre trop faible de sujets. Or plus un échantillon est important, plus grande est la précision dans l'estimation des résultats scientifiques. Dans le cas de la recherche animale, il est ainsi fréquent d'étudier l'efficacité de traitements de problèmes cardiaques sur moins de 10 animaux, groupe témoin inclus. Il est difficile d'imaginer comment des recherches sur un nombre de sujets aussi faible pourraient conduire à des résultats généralisables. Si, en psychologie, les recherches sur un échantillon aussi faible sont rares, la comparaison de groupes d'une vingtaine ou d'une trentaine de personnes a été la norme pendant des années. Certes, des groupes de cette taille peuvent être suffisants pour démontrer l'existence d'effets psychologiques extrêmement forts. L'effet *Stroop*, qui désigne la difficulté qu'éprouvent des sujets à prononcer un mot de couleur (par exemple, **rouge**) s'il est écrit dans une encre de couleur différente (par exemple, en **vert**) est un effet massif, extrêmement robuste et facilement démontrable avec un nombre très faible de participants. Mais toute démonstration rigoureuse d'effets plus subtils - et il s'agit probablement de la majorité des effets étudiés en psychologie - est tout simplement impossible avec des effectifs aussi réduits.

Pour éviter tous ces problèmes, j'ai essayé, dans les différentes expériences présentées ici, d'adopter certaines pratiques qui permettent de limiter le risque de faux positif, et de maximiser la puissance des expériences. Tout d'abord, presque toutes les expériences présentées ici font usage d'un plan expérimental croisé (*within-subject design*). Ce type d'expérience, où les sujets passent successivement les différentes conditions expérimentales, permet d'accentuer les chances de percevoir des différences entre différentes conditions expérimentales, si de telles différences existent. En effet, le fait de disposer des réponses des sujets pour toutes les conditions permet de minimiser le « bruit » et la variance liés aux différences inter-individuelles.

Deuxièmement, j'ai tenté de minimiser la flexibilité dans l'analyse en pré-enregistrant la moitié des expériences présentées dans cette thèse. Le pré-enregistrement consiste à écrire et

publier un plan d'analyse par avance, afin de se lier les mains et de limiter la tentation de faire des analyses expérimentales post-hoc.

Ensuite, de façon générale, je considère que les différences présentées ici sont robustes parce qu'elles montrent, dans la plupart des cas, des différences fortes entre conditions expérimentales, qui auraient eu très peu de chances d'émerger s'il n'y avait eu aucune différence entre les différentes conditions. En d'autres termes, plus techniques, l'ensemble des expériences présentées ici ont possédé des valeurs-p très basses, indiquant qu'un résultat extrêmement surprenant se serait produit si l'on faisait l'hypothèse d'une absence de différence. Dans ces conditions, les expériences présentées ici seraient probablement répliquables par une autre équipe universitaire.

Bibliographie

Lene AARØE et Michael Bang PETERSEN, « Crowding Out Culture: Scandinavians and Americans Agree on Social Welfare in the Face of Deservingness Cues », *The Journal of Politics*, 2014, vol. 76, n° 3, p. 684-697.

Arthur W. H. ADKINS, *Merit and Responsibility a Study in Greek Values*, Chicago, Univ of Chicago Press, 1975.

Christopher AKE, « Justice as Equality », *Philosophy & Public Affairs*, 1975, vol. 5, n° 1, p. 69-89.

George A. AKERLOF et Janet L. YELLEN, « Fairness and Unemployment », *The American Economic Review*, 1988, vol. 78, n° 2, p. 44-49.

George Arthur AKERLOF et Robert James SHILLER, *Les esprits animaux: comment nos comportements irrationnels gouvernent l'économie*, trad. fr. Corinne FAURE-GEORS, 2018.

Alberto ALESINA et George-Marios ANGELETOS, « Fairness and Redistribution: US vs. Europe », *American Economic Review*, 2005, vol. 95.

Joshua ALEXANDER, Ronald MALLON et Jonathan M. WEINBERG, « Accentuate the Negative », *Review of Philosophy and Psychology*, 2010, vol. 1, n° 2, p. 297–314.

Aurélien ALLARD, « La crise de la répliquabilité », *La Vie des idées*, 2018.

Aurélien ALLARD, « Mérite et travail ».

Elizabeth S. ANDERSON, « What Is the Point of Equality? », *Ethics*, 1999, vol. 109, n° 2, p. 287-337.

ARISTOTE, *Ethique à Nicomaque*, trad. fr. Jules TRICOT, Paris, Vrin, 1994.

Richard ARNESON, « Egalitarianism », in Edward N. ZALTA (dir.), *The Stanford Encyclopedia of Philosophy*, Summer 2013., Metaphysics Research Lab, Stanford University, 2013, p.

Richard ARNESON, « Luck Egalitarianism—A Primer », in Carl KNIGHT et Zofia STEMPOWSKA (dir.), *Responsibility and Distributive Justice*, Oxford University Press, 2011, p. 24–50.

Richard J. ARNESON, « Equality and Equal Opportunity for Welfare », *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 1989, vol. 56, n° 1, p. 77-93.

N. Scott ARNOLD, « Why Profits Are Deserved », *Ethics*, 1987, vol. 97, n° 2, p. 387-402.

Nomy ARPALY, *Merit, Meaning, and Human Bondage: An Essay on Free Will*, Princeton, N.J, Princeton University Press, 2006.

AUGUSTINUS, *La cité de Dieu*, Paris, Gallimard, 2008.

Carla BAGNOLI, « Constructivism in Metaethics », in Edward N. ZALTA (dir.), *The Stanford Encyclopedia of Philosophy*, Winter 2017., Metaphysics Research Lab, Stanford University, 2017, p.

Brian BARRY, *Political Argument: A Reissue With a New Introduction*, Berkeley, Univ of California Press, 1990.

Larry M. BARTELS, *Unequal Democracy: The Political Economy of the New Gilded Age*, New York, Princeton University Press, 2016.

Christian BAUDELLOT, Damien CARTRON, Jérôme GAUTIE et Olivier GODECHOT, *Bien ou mal payés ? : Les travailleurs du public et du privé jugent leurs salaires*, Paris, Rue d'Ulm, 2014.

Nicolas BAUMARD, *Comment nous sommes devenus moraux: une histoire naturelle du bien et du mal*, Paris, Jacob, 2010.

Nicolas BAUMARD, Jean-Baptiste ANDRÉ et Dan SPERBER, « A mutualistic approach to morality: The evolution of fairness by partner choice », *Behavioral and Brain Sciences*, 2013, vol. 36, n° 1, p. 59-78.

Nicolas BAUMARD et Coralie CHEVALLIER, « What Goes Around Comes Around: The Evolutionary Roots of the Belief in Immanent Justice », *Journal of Cognition and Culture*, 2012, vol. 12, n° 1-2, p. 67-80.

Nicolas BAUMARD, Olivier MASCARO et Coralie CHEVALLIER, « Preschoolers are able to take merit into account when distributing goods », *Developmental Psychology*, 2012, vol. 48, n° 2, p. 492-498.

Lawrence C. BECKER, « The Obligation to Work », *Ethics*, 1980, vol. 91, n° 1, p. 35-49.

Matthew S. BEDKE, « Intuitional Epistemology in Ethics », *Philosophy Compass*, 2010, vol. 5, n° 12, p. 1069-1083.

Laurent BÈGUE et Marina BASTOUNIS, « Two Spheres of Belief in Justice: Extensive Support for the Bidimensional Model of Belief in a Just World », *Journal of Personality*, 2003, vol. 71, n° 3, p. 435-463.

Roland BÉNABOU et Jean TIROLE, « Belief in a Just World and Redistributive Politics », *The Quarterly Journal of Economics*, 2006, vol. 121, n° 2, p. 699-746.

John J. BERMAN, Virginia MURPHY-BERMAN et Purnima SINGH, « Cross-Cultural Similarities and Differences in Perceptions of Fairness », *Journal of Cross-Cultural Psychology*, 1985, vol. 16, n° 1, p. 55-67.

Mitchell BERMAN, « Modest Retributivism », *Faculty Scholarship*, 2016.

Kenworthy BILZ, « Testing the Expressive Theory of Punishment », *Journal of Empirical Legal Studies*, 2016, vol. 13, n° 2, p. 358-392.

Christophe BOESCH et Hedwige BOESCH-ACHERMANN, *The chimpanzees of the Tai Forest: behavioural ecology and evolution*, Oxford, Oxford Univ. Press, 2008.

Pierre BOURDIEU et Jean-Claude PASSERON, *La reproduction: éléments pour une théorie du système d'enseignement*, Paris, Éd. de Minuit, 2005.

Pierre BOURDIEU et Jean-Claude PASSERON, *Les héritiers: les étudiants et la culture*, Paris, Les Éd. de Minuit, 1994.

Robert BOYD, Herbert GINTIS et Samuel BOWLES, « Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare », *Science*, 2010, vol. 328, n° 5978, p. 617-620.

John BRAITHWAITE et Philip PETTIT, *Not Just Deserts: A Republican Theory of Criminal Justice*, Oxford, Clarendon Press, 1993.

Bertolt BRECHT, *Sainte Jeanne des abattoirs*, Paris, L'Arche Editeur, 2008.

Bertolt BRECHT, *La Bonne Ame du Se-Tchouan*, trad. fr. Marie-Paule RAMO, Paris, L'Arche éditeur, 2004.

Bertolt BRECHT, *La vie de Galilée*, Paris, L'Arche, 1997.

Jessica BREGANT, Eugene M. CARUSO et Alex SHAW, *Expressing What? Evaluating the Expressive Value of Punishment*, Rochester, NY, Social Science Research Network, 2016.

Kimberley BROWNLEE, « 'Don't call people "rapists" : on the social contribution injustice of punishment' », *Current Legal Problems*, 2016, vol. 69, n° 1, p. 327-352.

Edmund BURKE, *Réflexions sur la Révolution en France*, Paris, Les Belles lettres, 2016.

John CARSON, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750-1940*, Princeton University Press, 2006.

Chris CHAMBERS, *The Seven Deadly Sins of Psychology: A Manifesto for Reforming the Culture of Scientific Practice*, Princeton University Press, 2017.

Shereen Jehan CHAUDHRY et George LOEWENSTEIN, « Thanking, Apologizing, Bragging, and Blaming: Responsibility Exchange Theory and the Currency of Communication », *Psychological Review*, à paraître.

Chao C. CHEN, « New Trends in Rewards Allocation Preferences: A Sino-U.S. Comparison », *Academy of Management Journal*, 1995, vol. 38, n° 2, p. 408-428.

Raj CHETTY, Nathaniel HENDREN, Patrick KLINE, Emmanuel SAEZ et Nicholas TURNER, « Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility », *The American Economic Review*, 2014, vol. 104, n° 5, p. 141-147.

Coralie CHEVALLIER, Jing XU, Kuniko ADACHI, Jean-Baptiste van der HENST et Nicolas BAUMARD, « Preschoolers' Understanding of Merit in Two Asian Societies », *PLOS ONE*, 2015, vol. 10, n° 5, p. e0114717.

Thomas CHRISTIANO, « The Tension between the Nature and the Norm of Voluntary Exchange », *The Southern Journal of Philosophy*, 2016, vol. 54, p. 109-129.

G. A. COHEN, *On the Currency of Egalitarian Justice, and Other Essays in Political Philosophy*, Princeton, Princeton University Press, 2011.

G. A. COHEN, *If You're an Egalitarian, How Come You're so Rich?*, Harvard University Press, 2001.

G. A. COHEN, « Reason, humanity, and the moral law », in Onora O'NEILL (dir.), *The Sources of Normativity*, Cambridge ; New York, Cambridge University Press, 1996, p. 167-188.

David COLAÇO, Markus KNEER, Joshua ALEXANDER et Edouard MACHERY, « On second thought: A refutation of the reflection defense », *manuscrit non publié*.

Alexander COPPOCK, « Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach », *Political Science Research and Methods*, 2018, p. 1-16.

J. Angelo CORLETT, « Making Sense of Retributivism », *Philosophy*, 2001, vol. 76, n° 1, p. 77-110.

Florian COVA (dir.), *La philosophie expérimentale*, Vuibert, 2012.

Florian COVA, *Qu'en pensez-vous? Introduction à la philosophie expérimentale*, Meaux, Germina, 2011.

Geoff CUMMING, *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis.*, New York, Routledge, 2011.

Geoffrey CUPIT, *Justice As Fittingness*, Oxford, Oxford University Press, 1997.

Fiery CUSHMAN, Anna DREBER, Ying WANG et Jay COSTA, « Accidental Outcomes Guide Punishment in a "Trembling Hand" Game », *PLoS ONE*, 2009, vol. 4, n° 8.

Jason DANA, George LOEWENSTEIN et Roberto WEBER, « Ethical immunity: How people violate their own moral standards without feeling they are doing so », in David De CREMER et Ann E. TENBRUNSEL (dir.), *Behavioral Business Ethics: Shaping an Emerging Field*, Routledge, 2012, p. 201-219.

Norman DANIELS, « Wide Reflective Equilibrium and Theory Acceptance in Ethics », *Journal of Philosophy*, 1979, vol. 76, n° 5, p. 256–282.

John M. DARLEY, Kevin M. CARLSMITH et Paul H. ROBINSON, « Incapacitation and Just Deserts as Motives for Punishment », *Law and Human Behavior*, 2000, vol. 24, n° 6, p. 659-683.

John M. DARLEY et Thane S. PITTMAN, « The Psychology of Compensatory and Retributive Justice », *Personality and Social Psychology Review*, 2003, vol. 7, n° 4, p. 324-336.

Ilan DAR-NIMROD et Steven J. HEINE, « Genetic Essentialism: On the Deceptive Determinism of DNA », *Psychological Bulletin*, 2011, vol. 137, n° 5, p. 800-818.

Shai DAVIDAI et Thomas GILOVICH, « Building a More Mobile America—One Income Quintile at a Time », *Perspectives on Psychological Science*, 2015, vol. 10, n° 1, p. 60-71.

Morton DEUTSCH, *Distributive Justice: A Social-Psychological Perspective*, New Haven, Yale Univ Press, 1985.

John DEWEY, *The collected works of John Dewey. 1932: Ethics*, Carbondale, Southern Illinois Univ. Press, 2008.

Ned DOBOS, « Exploitation, Working Poverty, and the Expressive Power of Wages », *Journal of Applied Philosophy*, à paraître.

John M. DORIS, *Lack of character: personality and moral behavior*, Cambridge, Cambridge Univ. Press, 2008.

Francois DUBET, *Injustices. L'expérience des inégalités au travail*, Paris, Points, 2013.

Francois DUBET, *Les places et les chances : Repenser la justice sociale*, Paris, Le Seuil, 2010.

R. A. DUFF, *Punishment, Communication, and Community*, OUP, 2003.

Emile DURKHEIM, *De la division du travail social*, Paris, Presses Universitaires de France - PUF, 2013.

Marie DURU-BELLAT, *Le mérite contre la justice*, Les Presses de Sciences Po, 2009.

Marie DURU-BELLAT, *Les inégalités sociales à l'école : genèse et mythes*, Paris, Presses Universitaires de France, 2003.

Ronald DWORKIN, *Sovereign Virtue: The Theory and Practice of Equality*, Harvard University Press, 2002.

Gideon ELFORD, « Social Class, Merit and Equality of Opportunity in Education », *Res Publica*, 2016, vol. 22, n° 3, p. 267-284.

Jon ELSTER, *Local Justice: How Institutions Allocate Scarce Goods and Necessary Burdens*, Russell Sage Foundation, 1993.

Birte ENGLISH, Thomas MUSSWEILER et Fritz STRACK, « Playing dice with criminal sentences: the influence of irrelevant anchors on experts' judicial decision making », *Personality & Social Psychology Bulletin*, 2006, vol. 32, n° 2, p. 188-200.

David ENOCH, *Taking morality seriously: a defense of robust realism*, Oxford, Oxford Univ. Press, 2013.

David ENOCH, « Ideal Theory, Utopianism, and What's the Question (in Political Philosophy) ».

Eva ERMAN et Niklas MÖLLER, « Practices and Principles: On the Methodological Turn in Political Theory », *Philosophy Compass*, 2015, vol. 10, n° 8, p. 533-546.

David M. ESTLUND, *Democratic authority: a philosophical framework*, Princeton, NJ, Princeton Univ. Press, 2009.

Lisa FARWELL et Bernard WEINER, « Bleeding Hearts and the Heartless: Popular Perceptions of Liberal and Conservative Ideologies », *Personality and Social Psychology Bulletin*, 2000, vol. 26, n° 7, p. 845-852.

Rainer Werner FASSBINDER, *Huit heures ne font pas un jour*, Carlotta Films.

Didier FASSIN, *L'Ombre du monde. Une anthropologie de la condition carcérale*, Points, 2017.

Didier FASSIN, *Punir. Une passion contemporaine*, Paris, Le Seuil, 2017.

N. T. FEATHER, « Distinguishing between deservingness and entitlement: earned outcomes versus lawful outcomes », *European Journal of Social Psychology*, 2003, vol. 33, n° 3, p. 367-385.

N. T. FEATHER, « Judgments of deservingness: studies in the psychology of justice and achievement », *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, 1999, vol. 3, n° 2, p. 86-107.

Joel FEINBERG, *Doing and Deserving: Essays in the Theory of Responsibility*, Princeton, Princeton University Press, 1974.

Joel FEINBERG, *Social philosophy*, Englewood Cliffs, NJ, Prentice Hall, 1973.

Joel FEINBERG, « The Expressive Function of Punishment », *The Monist*, 1965, vol. 49, n° 3, p. 397-423.

Fred FELDMAN, *Distributive Justice: Getting What We Deserve From Our Country*, Oxford, OUP Oxford, 2016.

Fred FELDMAN, *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*, Cambridge, Cambridge University Press, 1997.

Fred FELDMAN et Brad SKOW, « Desert », in Edward N. ZALTA (dir.), *The Stanford Encyclopedia of Philosophy*, Winter 2016., Metaphysics Research Lab, Stanford University, 2016, p.

John Martin FISCHER, *Deep Control: Essays on Free Will and Value*, Oxford, OUP USA, 2013.

John Martin FISCHER, Robert KANE, Derk PEREBOOM et Manuel VARGAS, *Four Views on Free Will*, Malden, MA, Wiley-Blackwell, 2007.

John Martin FISCHER et Mark RAVIZZA, *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, 1998.

Ronald FISCHER et Peter B. SMITH, « Reward Allocation and Culture: A Meta-Analysis », *Journal of Cross-Cultural Psychology*, 2003, vol. 34, n° 3, p. 251-268.

Philippa FOOT, « The Problem of Abortion and the Doctrine of Double Effect », *Oxford Review*, 1967, vol. 5, p. 5–15.

Michel FORSE, « Une passion de l'égalité particulièrement insatiable en France », *Revue de l'OFCE*, 2013, n° 131, p. 5-18.

Michel FORSE et Maxime PARODI, « La perception des inégalités en France depuis dix ans », *Revue de l'OFCE*, 2011, n° 118, p. 5-32.

Michel FORSE et Maxime PARODI, « Justice distributive », *Revue de l'OFCE*, 2006, no 98, n° 3, p. 213-244.

Harry G. FRANKFURT, *The Importance of What We Care About: Philosophical Essays*, Cambridge University Press, 1988.

Christopher FREIMAN et Shaun NICHOLS, « Is Desert in the Details? », *Philosophy and Phenomenological Research*, 2011, vol. 82, n° 1, p. 121-133.

Norman FROHLICH et Joe A. OPPENHEIMER, *Choosing Justice: An Experimental Approach to Ethical Theory*, Berkeley, University of California Press, 1992.

Norman FROHLICH, Joe A. OPPENHEIMER et Cheryl L. EAVEY, « Choices of Principles of Distributive Justice in Experimental Groups », *American Journal of Political Science*, 1987, vol. 31, n° 3, p. 606-636.

F. FUNK, V. MCGEER et M. GOLLWITZER, « Get the Message: Punishment Is Satisfying If the Transgressor Responds to Its Communicative Intent », *Personality and Social Psychology Bulletin*, 2014, vol. 40, n° 8, p. 986-997.

Raghavendra GADAGKAR, *Survival strategies: cooperation and conflict in animal societies*, Harvard University Press, 2001.

William A. GALSTON, *Justice and the Human Good*, Chicago, University of Chicago Press, 1980.

Allan GIBBARD, *Wise choices, apt feelings: a theory of normative judgment*, Oxford, Clarendon Press, 1990.

Gerd GIGERENZER et Peter M. TODD (dir.), *Simple heuristics that make us smart*, Oxford, Oxford Univ. Press, 2001.

Vladimir GIMPELSON et Daniel TREISMAN, « Misperceiving inequality », *Economics & Politics*, 2018, vol. 30, n° 1, p. 27-54.

Sharad GOEL, Adam OBENG et David ROTHSCHILD, « Non-Representative Surveys: Fast, Cheap, and Mostly Accurate », *manuscrit non publié*.

Deirdre GOLASH, *The Case Against Punishment: Retribution, Crime Prevention, and the Law*, New York, NYU Press, 2006.

M. GOLLWITZER et B. J. BUSHMAN, « Do Victims of Injustice Punish to Improve Their Mood? », *Social Psychological and Personality Science*, 2012, vol. 3, n° 5, p. 572-580.

Mario GOLLWITZER et Markus DENZLER, « What Makes Revenge Satisfactory: Seeing the Offender Suffer or Delivering a Message? », *Journal of Experimental Social Psychology*, 2009, vol. 45, n° 4, p. 840-844.

Mario GOLLWITZER et Milena MEDER, « What gives victims satisfaction when they seek revenge? », *European Journal of Social Psychology*, 2011, vol. 41, n° 3, p. 364-374.

Geoffrey P. GOODWIN et Adam BENFORADO, « Judging the Goring Ox: Retribution Directed Toward Animals », *Cognitive Science*, 2015, vol. 39, n° 3, p. 619-646.

Geoffrey P. GOODWIN et Justin F. LANDY, « Valuing different human lives », *Journal of Experimental Psychology. General*, 2014, vol. 143, n° 2, p. 778-803.

Daniela GOYA-TOCCHETTO, Matthew ECHOLS et Jen WRIGHT, « The lottery of life and moral desert: An empirical investigation », *Philosophical Psychology*, 2016, vol. 29, n° 8, p. 1112-1127.

W. I. GRIFFITH, « The effect of inputs and performance outcomes on allocation decisions », *Social Justice Research*, 1990, vol. 4, n° 1, p. 35-47.

Fred GROSKIND, « Public Reactions to Poor Families: Characteristics That Influence Attitudes toward Assistance », *Social Work*, 1991, vol. 36, n° 5, p. 446-453.

Madeleine GUILBERT, *Les Fonctions des femmes dans l'industrie*, Mouton et Cie, 1966.

Bertrand GUILLARME, *Penser la peine*, Paris, Presses Universitaires France, 2003.

GUILLERMINA JASSO et EVA M. MEYERSSON MILGROM, « Distributive Justice and CEO Compensation », *Acta Sociologica*, 2008, vol. 51, n° 2, p. 123-143.

Susan HAACK, *Evidence and inquiry: a pragmatist reconstruction of epistemology*, Prometheus Books, 2009.

Carolyn L. HAFER et Laurent BÈGUE, « Experimental Research on Just-World Theory: Problems, Developments, and Future Challenges. », *Psychological Bulletin*, 2005, vol. 131, n° 1, p. 128-167.

Jonathan HAIDT, « The emotional dog and its rational tail: A social intuitionist approach to moral judgment. », *Psychological Review*, 2001, vol. 108, n° 4, p. 814-834.

Jonathan HAIDT, Fredrik BJÖRKLUND et Scott MURPHY, « Moral Dumbfounding: When Intuition Finds No Reason », *manuscrit non publié*, p.

Jean HAMPTON, « Correction Harms versus Righting Wrongs: The Goal of Retribution », *UCLA Law Review*, 1991, vol. 39, p. 1659-1702.

Jean HAMPTON, « The Moral Education Theory of Punishment », *Philosophy & Public Affairs*, 1984, vol. 13, n° 3, p. 208-238.

Nathan HANNA, « Two Claims about Desert », *Pacific Philosophical Quarterly*, 2013, vol. 94, n° 1, p. 41-56.

Nathan HANNA, « The Passions of Punishment », *Pacific Philosophical Quarterly*, 2009, vol. 90, n° 2, p. 232-250.

Nathan HANNA, « Say What? A Critique of Expressive Retributivism », *Law and Philosophy*, 2008, vol. 27, n° 2, p. 123-150.

Robert D. HARE, *Without conscience: the disturbing world of the psychopaths among us*, New York, Guilford Press, 1999.

Sam HARRIS, *Moral landscape: how science can determine human values*, New York, Free Press, 2011.

H. L. A. HART, *Punishment and Responsibility: Essays in the Philosophy of Law*, Oxford, Oxford University Press, 2008.

H. L. A. HART, *Law, Liberty and Morality*, Oxford University Press, Oxford, 1968.

Nick HASLAM, Brock BASTIAN, Paul BAIN et Yoshihisa KASHIMA, « Psychological Essentialism, Implicit Theories, and Intergroup Relations », *Group Processes & Intergroup Relations*, 2006, vol. 9, n° 1, p. 63-76.

Nick HASLAM, Louis ROTHSCHILD et Donald ERNST, « Essentialist beliefs about social categories », *British Journal of Social Psychology*, 2000, vol. 39, n° 1, p. 113-127.

Albert H. HASTORF et Hadley CANTRIL, « They saw a game; a case study », *The Journal of Abnormal and Social Psychology*, 1954, vol. 49, n° 1, p. 129-134.

F. A. HAYEK, *Law, Legislation and Liberty*, London, Routledge, 2012.

F. A. HAYEK, *The Constitution of Liberty*, Chicago, University of Chicago Press, 2011.

Andrew von HIRSCH, *Censure and Sanctions*, Oxford, Oxford University Press, 1994.

Jennifer L. HOCHSCHILD, *What's Fair: American Beliefs about Distributive Justice*, Cambridge, Mass., Harvard University Press, 1986.

Nils HOLTUG et Kasper LIPPERT-RASMUSSEN (dir.), *Egalitarianism: New Essays on the Nature and Value of Equality*, Oxford, Clarendon Press, 2007.

Axel HONNETH, *Le droit de la liberté: Esquisse d'une éthicité démocratique*, trad. fr. Frédéric JOLY et trad. fr. Pierre RUSCH, Paris, Gallimard, 2015.

Axel HONNETH, *La société du mépris : vers une nouvelle théorie critique*, trad. fr. Olivier VOIROL, Paris, La Découverte, 2015.

Axel HONNETH, *La lutte pour la reconnaissance*, trad. fr. Pierre RUSCH, Paris, Folio, 2013.

David W. HOWELLS, Emily S. SENA et Malcolm R. MACLEOD, « Bringing rigour to translational medicine », *Nature Reviews Neurology*, 2014, vol. 10, n° 1, p. 37-43.

Michael HUEMER, « Phenomenal Conservatism and the Internalist Intuition », *American Philosophical Quarterly*, 2006, vol. 43, n° 2, p. 147-158.

David HUME, *A Treatise of Human Nature*, Mineola, N.Y, Dover Publications, 2003.

David HUME, *An Enquiry Concerning the Principles of Morals*, Indianapolis, Hackett Publishing, 1983.

Thomas HURKA, « The Common Structure of Virtue and Desert », *Ethics*, 2001, vol. 112, n° 1, p. 6-31.

Susan L. HURLEY, *Justice, Luck and Knowledge*, Harvard University Press, 2005.

Douglas N. HUSAK, « Why Punish the Deserving? », *Noûs*, 1992, vol. 26, n° 4, p. 447-464.

Ronnie JANOFF-BULMAN, « Characterological versus behavioral self-blame: Inquiries into depression and rape », *Journal of Personality and Social Psychology*, 1979, vol. 37, n° 10, p. 1798-1809.

Agnieszka JAWORSKA et Julie TANNENBAUM, « The Grounds of Moral Status », in Edward N. ZALTA (dir.), *The Stanford Encyclopedia of Philosophy*, Spring 2018., Metaphysics Research Lab, Stanford University, 2018, p.

JOHN J. BERMAN, VIRGINIA MURPHY-BERMAN et PURNIMA SINGH, « Cross-Cultural Similarities and Differences in Perceptions of Fairness », *Journal of Cross-Cultural Psychology*, 1985, vol. 16, n° 1, p. 55-67.

John T. JOST, Mahzarin R. BANAJI et Brian A. NOSEK, « A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo », *Political Psychology*, 2004, vol. 25, n° 6, p. 881-919.

John T. JOST et Orsolya HUNYADY, « Antecedents and Consequences of System-Justifying Ideologies », *Current Directions in Psychological Science*, 2005, vol. 14, n° 5, p. 260-265.

Richard JOYCE, *The evolution of morality*, Cambridge, Mass., The MIT Press, 2007.

Shelly KAGAN, *The Geometry of Desert*, New York, Oxford University Press, 2015.

Daniel KAHNEMAN et Gary KLEIN, « Conditions for intuitive expertise: a failure to disagree », *The American Psychologist*, 2009, vol. 64, n° 6, p. 515-526.

Daniel KAHNEMAN, Jack L. KNETSCH et Richard THALER, « Fairness as a Constraint on Profit Seeking: Entitlements in the Market », *The American Economic Review*, 1986, vol. 76, n° 4, p. 728-741.

Daniel KAHNEMAN, David SCHKADE et Cass SUNSTEIN, « Shared Outrage and Erratic Awards: The Psychology of Punitive Damages », *Journal of Risk and Uncertainty*, 1998, vol. 16, n° 1, p. 49-86.

Daniel KAHNEMAN et Cass SUNSTEIN, « Indignation: Psychology, Politics, Law », *Law & Economics Working Papers*, 2007.

Jocelyn KAISER, « Plan to replicate 50 high-impact cancer papers shrinks to just 18 », *Science / AAAS*, 2018, 31/07/2018.

Robert KANE, *Free Will and Values*, Albany, State University of New York Press, 1985.

Emmanuel KANT, *Métaphysique des moeurs : Tome 2, Doctrine du droit ; Doctrine de la vertu*, Flammarion, 2018.

Immanuel KANT, *Fondements de la métaphysique des moeurs*, trad. fr. Victor DELBOS, Paris, Librairie Générale Française, 1993.

Michael B. KATZ, *The Undeserving Poor: America's Enduring Confrontation with Poverty: Fully Updated and Revised*, Oxford, Oxford University Press, 2013.

Egon KAYSER et Helmut LAMM, « Input integration and input weighting in decisions on allocations of gains and losses », *European Journal of Social Psychology*, 1980, vol. 10, n° 1, p. 1-15.

Johannes KELLER, « In genes we trust: the biological component of psychological essentialism and its relationship to mechanisms of motivated social cognition », *Journal of Personality and Social Psychology*, 2005, vol. 88, n° 4, p. 686-702.

Stephen KERSHNER, *Desert and Virtue: A Theory of Intrinsic Value*, Lanham, Lexington Books, 2009.

Philip KITCHER, *The ethical project*, Cambridge, Mass., Harvard Univ. Press, 2014.

Richard A. KLEIN, Kate A. RATLIFF, Michelangelo VIANELLO, Reginald B. ADAMS JR., Štěpán BAHNÍK, Michael J. BERNSTEIN, ... et Brian A. NOSEK, « Investigating variation in replicability: A “many labs” replication project », *Social Psychology*, 2014, vol. 45, n° 3, p. 142-152.

Richard Anthony KLEIN, Michelangelo VIANELLO, Fred HASSELMAN, Byron Gregory ADAMS, Reginald B. ADAMS, Sinan ALPER, ... et Brian A. NOSEK, « Many Labs 2: Investigating Variation in Replicability Across Sample and Setting ».

Carl KNIGHT, « Luck Egalitarianism », *Philosophy Compass*, 2013, vol. 8, n° 10, p. 924-934.

Carl KNIGHT et Zofia STEMPOWSKA (dir.), *Responsibility and Distributive Justice*, Oxford University Press, 2014.

Joshua KNOBE, *Experimental Philosophy*, OUP USA, 2008.

Joshua KNOBE et Shaun NICHOLS, « An Experimental Philosophy Manifesto », in Joshua KNOBE et Shaun NICHOLS (dir.), *Experimental Philosophy*, Oxford University Press, 2007, p. 3-14.

Valerie S. KNOPIK, Jenae M. NEIDERHISER, John C. DEFRIES et Robert PLOMIN, *Behavioral Genetics*, 7^e édition., New York, Worth Publishers, 2016.

James KONOW, « Which Is the Fairest One of All? A Positive Analysis of Justice Theories », *Journal of Economic Literature*, 2003, vol. 41, n° 4, p. 1188-1239.

James KONOW, « Fair and square: the four sides of distributive justice », *Journal of Economic Behavior & Organization*, 2001, vol. 46, n° 2, p. 137-164.

Christine M. KORSGAARD, *The Sources of Normativity*, Cambridge ; New York, Cambridge University Press, 1996.

Ziva KUNDA, « The case for motivated reasoning. », *Psychological bulletin*, 1990, vol. 108, n° 3, p. 480-498.

Helmut LAMM, Egon KAYSER et Volker SCHANZ, « An Attributional Analysis of Interpersonal Justice: Ability and Effort as Inputs in the Allocation of Gain and Loss », *The Journal of Social Psychology*, 1983, vol. 119, n° 2, p. 269-281.

Hélène LANDEMORE et Jon ELSTER (dir.), *Collective wisdom: principles and mechanisms*, Cambridge, Cambridge Univ. Press, 2012.

Melvin J. LERNER, « The Two Forms of Belief in a Just World », in Leo MONTADA et Melvin J. LERNER (dir.), *Responses to Victimizations and Belief in a Just World*, Boston, MA, Springer US, 1998, p. 247-269.

Melvin J. LERNER, *The belief in a just world: a fundamental delusion*, New York, NY, Plenum Press, 1980.

Gerald S. LEVENTHAL et James W. MICHAELS, « Locus of cause and equity motivation as determinants of reward allocation », *Journal of Personality and Social Psychology*, 1971, vol. 17, n° 3, p. 229-235.

Neil LEVY et Michael MCKENNA, « Recent Work on Free Will and Moral Responsibility », *Philosophy Compass*, 2009, vol. 4, n° 1, p. 96-133.

P. LIÉNARD, C. CHEVALLIER, O. MASCARO, P. KIURA et N. BAUMARD, « Early Understanding of Merit in Turkana Children », *Journal of Cognition and Culture*, 2013, vol. 13, n° 1-2, p. 57-66.

Matthew LINDAUER et Christian BARRY, « Moral Judgment and the Duties of Innocent Beneficiaries of Injustice », *Review of Philosophy and Psychology*, 2017, vol. 8, n° 3, p. 671-686.

Robert LINHART, *L'établi*, Paris, Editions de Minuit, 1981.

Kasper LIPPERT-RASMUSSEN, « Justice and Bad Luck », in Edward N. ZALTA (dir.), *The Stanford Encyclopedia of Philosophy*, Summer 2018., Metaphysics Research Lab, Stanford University, 2018, p.

Kasper LIPPERT-RASMUSSEN, *Luck Egalitarianism*, London, Bloomsbury Academic, 2015.

Margit E. Oswald LIVIA B. KELLER, « A Closer Look at an Eye for an Eye: Laypersons' Punishment Decisions Are Primarily Driven by Retributive Motives », *Soc. Justice Res.*, 2010, vol. 23, n° 2, p. 99-116.

J. R. LUCAS, *On Justice*, Oxford, Oxford University Press, 1980.

Kirk LUDWIG, « The Epistemology of Thought Experiments: First Person versus Third Person Approaches », *Midwest Studies In Philosophy*, 2007, vol. 31, n° 1, p. 128-159.

Edouard MACHERY, *Philosophy Within Its Proper Bounds*, Oxford, United Kingdom, Oxford University Press, 2017.

Alasdair MACINTYRE, *After Virtue: A Study in Moral Theory, Third Edition*, Notre Dame, Ind, University of Notre Dame Press, 2007.

John L. MACKIE, *Ethics: inventing right and wrong*, London, Penguin Books, 1990.

Irwin MAHLER, Lisa GREENBERG et Haruo HAYASHI, « A comparative study of rules of justice: Japanese versus American », *Psychologia: An International Journal of Psychology in the Orient*, 1981, vol. 24, n° 1, p. 1-8.

Pierre MANENT (dir.), *Les libéraux. 2*, Paris, Hachette, 1986.

Francois MANIQUET et Marc FLEURBAEY, *Equality of Opportunity: The Economics of Responsibility*, New Jersey, World Scientific, 2012.

Justin W. MARTIN et Fiery CUSHMAN, « To Punish or to Leave: Distinct Cognitive Processes Underlie Partner Control and Partner Choice Behaviors », *PLOS ONE*, 2015, vol. 10, n° 4, p. e0125193.

Andrew MASON, *Levelling the Playing Field: The Idea of Equal Opportunity and Its Place in Egalitarian Thought*, New York, Oxford University Press, 2006.

Andrew MASON, « Equality of Opportunity, Old and New », *Ethics*, 2001, vol. 111, n° 4, p. 760-781.

Sarah MATHEW et Robert BOYD, « The cost of cowardice: punitive sentiments towards free riders in Turkana raids », *Evolution and Human Behavior*, 2014, vol. 35, n° 1, p. 58-64.

Matt MATRAVERS, *Justice and Punishment: The Rationale of Coercion*, OUP Oxford, 2000.

Robert MAYER, « What's Wrong with Exploitation? », *Journal of Applied Philosophy*, 2007, vol. 24, n° 2, p. 137-150.

Leslie MCCALL, *The Undeserving Rich: American Beliefs about Inequality, Opportunity, and Redistribution*, New York, NY, Cambridge University Press, 2013.

Herbert MCCLOSKEY et John ZALLER, *The American Ethos: Public Attitudes Toward Capitalism and Democracy*, Harvard University Press, 1984.

Robert M. MCFATTER, « Purposes of punishment: Effects of utilities of criminal sanctions on perceived appropriateness », *Journal of Applied Psychology*, 1982, vol. 67, n° 3, p. 255-267.

Owen MCLEOD, « Desert and Institutions », in Louis P. POJMAN et Owen MCLEOD (dir.), *What Do We Deserve?: A Reader on Justice and Desert*, Oxford University Press, 1999, p. 186.

Owen MCLEOD, « Desert and Wages », *Utilitas*, 1996, vol. 8, n° 2, p. 205-221.

J. Ellis MCTAGGART, « Hegel's Theory of Punishment », *International Journal of Ethics*, 1896, vol. 6, n° 4, p. 479-502.

Barbara A. MELLERS et Jonathan BARON (dir.), *Psychological perspectives on justice: theory and applications*, Cambridge, Cambridge Univ. Press, 1993.

- Yves MICHAUD, *Qu'est-ce que le mérite ?*, Paris, Folio, 2011.
- Philip A. MICHELBAUGH, John T. SCOTT, Richard E. MATLAND et Brian H. BORNSTEIN, « Doing Rawls Justice: An Experimental Study of Income Distribution Norms », *American Journal of Political Science*, 2003, vol. 47, n° 3, p. 523-539.
- John Stuart MILL, *L'utilitarisme*, trad. fr. Georges TANESSE, Paris, Flammarion, 2010.
- David MILLER, *Strangers in our midst: the political philosophy of immigration*, Harvard University Press, 2016.
- David MILLER, *Justice for Earthlings: Essays in Political Philosophy*, Cambridge, Cambridge University Press, 2013.
- David MILLER, *Principles of Social Justice*, Harvard University Press, 2001.
- David MILLER, « Complex Equality », in David MILLER et Michael WALZER (dir.), *Pluralism, Justice, and Equality*, Oxford University Press, 1995, p. 197-225.
- David MILLER, « Distributive Justice: What the People Think », *Ethics*, 1992, vol. 102, n° 3, p. 555-593.
- David MILLER, *Market, state, and community: theoretical foundations of market socialism*, Oxford, Clarendon, 1990.
- David MILLER, *Social Justice*, Oxford, Oxford University Press, 1979.
- Gregory MITCHELL, Philip E. TETLOCK, Barbara A. MELLERS et Lisa D. ORDÓÑEZ, « Judgments of Social Justice: Compromises Between Equality and Efficiency », *Journal of Personality and Social Psychology*, 1993, vol. 65, n° 4, p. 629-639.
- Michael S. MOORE, « Justifying Retributivism », *Israel Law Review*, 1993, vol. 27, n° 1-2, p. 15-49.
- Herbert MORRIS, « A Paternalistic Theory of Punishment », *American Philosophical Quarterly*, 1981, vol. 18, n° 4, p. 263-271.
- Herbert MORRIS, « Persons and Punishment », *The Monist*, 1968, vol. 52, n° 4, p. 475-501.
- Thomas MULLIGAN, *Justice and the Meritocratic State*, Routledge, 2017.
- Kevin J. MULLINIX, Thomas J. LEEPER, James N. DRUCKMAN et Jeremy FREESE, « The Generalizability of Survey Experiments », *Journal of Experimental Political Science*, 2015, vol. 2, n° 2, p. 109-138.
- Liam MURPHY et Thomas NAGEL, *The Myth of Ownership: Taxes and Justice*, Oxford, Oxford University Press, 2004.
- Vignia MURPHY-BERMAN, John J. BERMAN, Purnima SINGH, Anju PACHAURI et Pramod KUMAR, « Factors affecting allocation to needy and meritorious recipients: A cross-cultural comparison », *Journal of Personality and Social Psychology*, 1984, vol. 46, n° 6, p. 1267-1272.

Thomas NADELHOFFER, Saeideh HESHMATI, Deanna KAPLAN et Shaun NICHOLS, « Folk Retributivism and the Communication Confound », *Economics & Philosophy*, 2013, vol. 29, n° 2, p. 235-261.

T. NAGEL, « Justice and Nature », *Oxford Journal of Legal Studies*, 1997, vol. 17, n° 2, p. 303–322.

Thomas NAGEL, *Equality and Partiality*, Oxford University Press, 1995.

Uma NARAYAN, « Appropriate Responses and Preventive Benefits: Justifying Censure and Hard Treatment in Legal Punishment », *Oxford Journal of Legal Studies*, 1993, vol. 13, n° 2, p. 166-182.

Jules NAUDET, *Entrer dans l'élite : Parcours de réussite en France, aux Etats-Unis et en Inde*, Paris, Presses Universitaires de France - PUF, 2012.

Dana K. NELKIN, « Moral Luck », in Edward N. ZALTA (dir.), *The Stanford Encyclopedia of Philosophy*, Winter 2013., Metaphysics Research Lab, Stanford University, 2013.

Leif D. NELSON, Joseph SIMMONS et Uri SIMONSOHN, « Psychology's Renaissance », *Annual Review of Psychology*, 2018, vol. 69, n° 1.

Marc NEUBERG (dir.), *La Responsabilité : Questions philosophiques*, Paris, Presses Universitaires de France - PUF, 1997.

Marc NEUBERG (dir.), *Théorie de l'action*, Liège, Editions Mardaga, 1995.

Shaun NICHOLS, *Bound: Essays on Free Will and Responsibility*, Oxford University Press, USA, 2017.

Shaun NICHOLS, « The Indeterminist Intuition: Source and Status », *The Monist*, 2012, vol. 95, n° 2, p. 290-307.

Shaun NICHOLS, « Norms with feeling: towards a psychological account of moral judgment », *Cognition*, 2002, vol. 84, n° 2, p. 221-236.

R. S. NICKERSON, « Null hypothesis significance testing: a review of an old and continuing controversy », *Psychological Methods*, 2000, vol. 5, n° 2, p. 241-301.

Gregory B NORTHCRAFT et Margaret A NEALE, « Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions », *Organizational Behavior and Human Decision Processes*, 1987, vol. 39, n° 1, p. 84-97.

Michael I. NORTON et Dan ARIELY, « Building a Better America—One Wealth Quintile at a Time », *Perspectives on Psychological Science*, 2011, vol. 6, n° 1, p. 9-12.

Robert NOZICK, *Anarchy, State, and Utopia*, New York, NY, Basic Books, 2013.

Robert NOZICK, *Philosophical Explanations*, Cambridge, Mass., Harvard University Press, 1981.

Robert NOZICK, « Retributive punishment », in *Philosophical Explanations*, Cambridge, Mass., Harvard University Press, 1981, p. 363-398.

Serena OLSARETTI (dir.), *Desert and Justice*, Oxford, Clarendon Press, 2007.

Serena OLSARETTI, *Liberty, desert and the market: a philosophical study*, Cambridge, Cambridge Univ. Press, 2004.

Sidney ORREN et Gary VERBA, *Equality in America: A View from the Top by Sidney Verba*, Harvard University Press, 1985.

Derek PARFIT, *On What Matters: Volume One*, Oxford, Oxford University Press, 2013.

Derek PARFIT, *Reasons and Persons*, Oxford, Oxford University Press, 1986.

Philippe Van PARIJS, *Real Freedom for All*, Oxford, Clarendon Press, 1998.

Jean-Claude PASSERON et Pierre BOURDIEU, *La Reproduction*, Paris, Editions de Minuit, 1970.

Indrajeet PATIL, Micaela ZUCHELLI, Wouter KOOL, Stephanie CAMPBELL, Federico FORNASIER, Marta CALO, Giorgia SILANI, Mina CIKARA et Fiery Andrews CUSHMAN, « Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures », *manuscrit non publié*.

Serge PAUGAM et Marion SELZ, « La perception de la pauvreté en Europe depuis le milieu des années 1970. Analyse des variations structurelles et conjoncturelles », *Economie et Statistique*, 2005, vol. 383, n° 1, p. 283-305.

Charles S. PEIRCE, *Ecrits sur le signe*, Points, 2017.

Derk PEREBOOM, *Living without Free Will*, Cambridge University Press, 2001.

Thomas PIKETTY, *Le Capital au XXI^e siècle*, Paris, Le Seuil, 2013.

Steven PINKER, *The better angels of our nature*, London, Penguin, 2012.

Robert PLOMIN, *Blueprint: how DNA makes us who we are*, Cambridge, Massachusetts, The MIT Press, 2018.

Robert PLOMIN, John C. DEFRIES, Valerie S. KNOPIK et Jenae M. NEIDERHISER, « Top 10 Replicated Findings from Behavioral Genetics », *Perspectives on psychological science : a journal of the Association for Psychological Science*, 2016, vol. 11, n° 1, p. 3-23.

Louis P. POJMAN, « Are Human Rights Based on Equal Human Worth? », *Philosophy and Phenomenological Research*, 1992, vol. 52, n° 3, p. 605-622.

Louis P. POJMAN, « The Moral Status of Affirmative Action », *Public Affairs Quarterly*, 1992, vol. 6, n° 2, p. 181-206.

Louis P. POJMAN et Owen MCLEOD (dir.), *What Do We Deserve?: A Reader on Justice and Desert*, Oxford University Press, USA, 1998.

Igor PRIMORATZ, « Punishment as Language », *Philosophy*, 1989, vol. 64, n° 248, p. 187-205.

James RACHELS, « Punishment and Desert », in Hugh LAFOLLETTE - (dir.), *Ethics in Practice*, Blackwell, 1997, p. 466–74.

K. A. RASINSKI, « What's fair is fair--or is it? Value differences underlying public views about social justice », *Journal of Personality and Social Psychology*, 1987, vol. 53, n° 1, p. 201-211.

John RAWLS, *La justice comme équité*, La Découverte, 2008.

John RAWLS, *Political Liberalism*, New York, Columbia University Press, 2005.

John RAWLS, *Collected Papers*, Revised ed. edition., Cambridge, Mass., Harvard University Press, 2001.

John RAWLS, « Kantian Constructivism in Moral Theory », *The Journal of Philosophy*, 1980, vol. 77, n° 9, p. 515-572.

John RAWLS, « The Independence of Moral Theory », *Proceedings and Addresses of the American Philosophical Association*, 1974, vol. 48, p. 5-22.

John RAWLS et Catherine AUDARD, *Théorie de la justice*, Points Seuil, 2009.

Emmanuel RENAULT, *L'expérience de l'injustice: essai sur la théorie de la reconnaissance*, Paris, la Découverte, 2017.

Charles RENOUVIER, *Science de la morale*, Paris, Fayard, 2002.

Stanley REST, Richard NIERENBERG, Bernard WEINER et Heinz HECKHAUSEN, « Further evidence concerning the effects of perceptions of effort and ability on achievement evaluation. », *Journal of Personality and Social Psychology*, 1973, vol. 28, n° 2, p. 187-191.

Arthur RIPSTEIN, « Equality, Luck, and Responsibility », *Philosophy & Public Affairs*, 1994, vol. 23, n° 1, p. 3-23.

Paul H. ROBINSON et John M. DARLEY, « Does Criminal Law Deter? A Behavioural Science Investigation », *Oxford Journal of Legal Studies*, 2004, vol. 24, n° 2, p. 173-205.

Paul H. ROBINSON et John M. DARLEY, « Utility of Desert », *Northwestern University Law Review*, 1996, vol. 91, p. 453.

John E. ROEMER, *Equality of Opportunity*, Harvard University Press, 2000.

Pierre ROSANVALLON, *La Société des égaux*, Paris, Points, 2013.

David ROSS, *The Right and the Good*, Oxford, Clarendon Press, 2003.

Donald ROY, *Un sociologue à l'usine*, Paris, La Découverte, 2006.

Edward ROYZMAN, Kwanwoo KIM et Robert LEEMAN, « The curious tale of Julie and Mark: Unraveling the moral dumfounding effect », *Judgment and Decision Making*, 2015, vol. 10, p. 296-313.

Jesper RYBERG (dir.), *Popular punishment: on the normative significance of public opinion*, Oxford, Oxford Univ. Press, 2014.

Wojciech SADURSKI, *Giving Desert Its Due: Social Justice and Legal Theory*, Springer, 2010.

Jean SAGLIO, « Négociation de classifications et régulation salariale dans le système français de relations professionnelles », *Travail et Emploi*, 1988, N°38.

Jean SAGLIO, « Hiérarchies salariales et négociations de classifications - France, 1900-1950 », *Travail et Emploi*, 1986, N°27.

Michael J. SANDEL, *Liberalism and the Limits of Justice*, Cambridge, Cambridge University Press, 1998.

Jean-Paul SARTRE, *Les mots*, Paris, Gallimard, 2009.

Patrick SAVIDAN, *Repenser l'égalité des chances*, Paris, Hachette Littératures, 2010.

T. M. SCANLON, *What We Owe to Each Other*, Cambridge, Mass., Belknap Press, 2000.

T. M. SCANLON, « The Significance of Choice », in Sterling M. McMURRIN (dir.), *The Tanner Lectures on Human Values (Vol. 8, pp. 149-216)*, University of Utah Press, 1988, p.

Thomas SCANLON, *Why does inequality matter?*, Oxford New York, Oxford University Press, 2018.

Samuel SCHEFFLER, « Distributive justice and economic desert », in Serena OLSARETTI (dir.), *Desert and Justice*, Oxford, Clarendon Press, 2007, p.

Samuel SCHEFFLER, « What Is Egalitarianism? », *Philosophy & Public Affairs*, 2003, vol. 31, n° 1, p. 5-39.

Samuel SCHEFFLER, « Responsibility, Reactive Attitudes, and Liberalism in Philosophy and Politics », *Philosophy and Public Affairs*, 1992, n° 21, p. 299-323.

David SCHMIDTZ, *The Elements of Justice*, Cambridge, Cambridge University Press, 2006.

David SCHMIDTZ et Robert E. GOODIN, *Social Welfare and Individual Responsibility*, Cambridge, Cambridge University Press, 1998.

Eric SCHWITZGEBEL et Fiery CUSHMAN, « Philosophers' biased judgments persist despite training, expertise and reflection », *Cognition*, 2015, vol. 141, p. 127-137.

Eric SCHWITZGEBEL et Fiery CUSHMAN, « Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers », *Mind & Language*, 2012, vol. 27, n° 2, p. 135-153.

John T. SCOTT, Richard E. MATLAND, Philip A. MICHELBAACH et Brian H. BORNSTEIN, « Just Deserts: An Experimental Study of Distributive Justice Norms », *American Journal of Political Science*, 2001, vol. 45, n° 4, p. 749-767.

Shlomi SEGALL, *Why Inequality Matters: Luck Egalitarianism, its Meaning and Value*, Cambridge University Press, 2016.

Amartya SEN, *Development as freedom*, Oxford, Univ. Press, 2001.

Russ SHAFER-LANDAU, *Moral realism: a defence*, Oxford, Clarendon Press, 2005.

Russ SHAFER-LANDAU, « The Failure of Retributivism », *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 1996, vol. 82, n° 3, p. 289-316.

Russ SHAFER-LANDAU, « Can Punishment Morally Educate? », *Law and Philosophy*, 1991, vol. 10, n° 2, p. 189-219.

Nichols SHAUN, « Brute Retributivism », in Thomas A. NADELHOFFER (dir.), *The Future of Punishment*, Oxford University Press, 2013, p. 25-46.

George SHER, *In Praise of Blame*, Oxford, Oxford University Press, 2007.

George SHER, *Desert*, Princeton, NJ, Princeton University Press, 1989.

Henry SIDGWICK, *The Methods of Ethics*, Indianapolis, Hackett Publishing Company, 1981.

Joseph P. SIMMONS, Leif D. NELSON, Jeff GALAK et Shane FREDERICK, « Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds », *Journal of Consumer Research*, 2011, vol. 38, n° 1, p. 1-15.

Arne SJÖSTRÖM et Mario GOLLWITZER, « Displaced revenge: Can revenge taste “sweet” if it aims at a different target? », *Journal of Experimental Social Psychology*, 2015, vol. 56.

A. J. SKILLEN, « How to Say Things with Walls », *Philosophy*, 1980, vol. 55, n° 214, p. 509-523.

Saul SMILANSKY, « Control, Desert and the Difference between Distributive and Retributive Justice », *Philosophical Studies*, 2006, vol. 131, n° 3, p. 511-524.

Saul SMILANSKY, *Free Will and Illusion*, Oxford: New York, Oxford University Press, 2000.

Adam SMITH, *Théorie des sentiments moraux*, PUF, 2014.

Adam SMITH, *The Theory of Moral Sentiments*, New York, N.Y, Penguin Classics, 2010.

Eliot R. SMITH et James R. KLUEGEL, *Beliefs About Inequality: Americans' Views of What Is and What Ought to Be*, New York, Aldine Transaction, 1986.

Karol Edward SOŁTAN, *The causal theory of justice*, Berkeley, Univ. of California Press, 1987.

Tamler SOMMERS, *Relative Justice: Cultural Diversity, Free Will, and Moral Responsibility*, Princeton, N.J, Princeton University Press, 2012.

Tamler SOMMERS, « Partial Desert », in David SHOEMAKER (dir.), *Oxford Studies in Agency and Responsibility*, Oxford University Press, forthcoming, p.

Jean-Fabien SPITZ, *Abolir le hasard? Responsabilité individuelle et justice sociale*, Paris, Librairie Philosophique Vrin, 2008.

Matthew L. STANLEY, Ashley M. DOUGHERTY, Brenda W. YANG, Paul HENNE et Felipe DE BRIGARD, « Reasons probably won't change your mind: The role of reasons in revising moral decisions », *Journal of Experimental Psychology: General*, 2018, vol. 147, n° 7, p. 962-987.

Christina STARMANS, Mark SHESKIN et Paul BLOOM, « Why people prefer unequal societies », *Nature Human Behaviour*, 2017, vol. 1, n° 4, p. 0082.

Allyn Maclean STEARMAN, *Yuqui: Forest Nomads in a Changing World*, Holt, Rinehart and Winston, 1989.

Hillel STEINER, « Debate: Levels of Non-ideality », *Journal of Political Philosophy*, 2017, vol. 25, n° 3, p. 376-384.

Galen J. STRAWSON, « The Impossibility of Moral Responsibility », *Philosophical Studies*, 1994, vol. 75, n° 1-2, p. 5-24.

Sharon STREET, « Coming to Terms with Contingency: Humean Constructivism About Practical Reason », in Jimmy LENMAN et Yonatan SHEMMER (dir.), *Constructivism in Practical Philosophy*, Oxford University Press, 2012, p.

Sharon STREET, « A Darwinian Dilemma for Realist Theories of Value », *Philosophical Studies*, 2006, vol. 127, n° 1, p. 109-166.

Kok-Chor TAN, *Justice, Institutions, and Luck: The Site, Ground, and Scope of Equality*, Oxford, Oxford University Press, 2014.

John TASIIOULAS, « Punishment and Repentance », *Philosophy*, 2006, vol. 81, n° 2, p. 279-322.

Charles TAYLOR, *The ethics of authenticity*, Harvard University Press, 2018.

Larry TEMKIN, « Justice, Equality, Fairness, Desert, Rights, Free Will, Responsibility, and Luck », in Carl KNIGHT et Zofia STEMPLOWSKA (dir.), *Responsibility and Distributive Justice*, Oxford University Press, 2011, p.

Larry S. TEMKIN, *Inequality*, New York, Oxford Univ. Press, 1996.

Elise TENRET, *Les étudiants et le mérite - À quoi bon être diplômé ?*, Paris, La Documentation Française, 2011.

THOMAS, *Textes sur la morale*, Paris, J. Vrin, 2011.

Judith Jarvis THOMSON, « Killing, Letting Die, and the Trolley Problem », *The Monist*, 1976, vol. 59, n° 2, p. 204-217.

Michael TOMASELLO, *Why we cooperate*, Cambridge, Mass., MIT Press, 2009.

Michael TONRY (dir.), *Retributivism Has a Past: Has It a Future?*, Oxford, Oxford University Press, 2011.

Robert L. TRIVERS, « The Evolution of Reciprocal Altruism », *The Quarterly Review of Biology*, 1971, vol. 46, n° 1, p. 35-57.

Eric TURKHEIMER, « Three Laws of Behavior Genetics and What They Mean », *Current Directions in Psychological Science*, 2000, vol. 9, n° 5, p. 160-164.

Laura VALENTINI, « Social Samaritan Justice: When and Why Needy Fellow Citizens Have a Right to Assistance », *American Political Science Review*, 2015, vol. 109, n° 4, p. 735-749.

Peter VALLENTYNE, « Brute luck and responsibility », *Politics, Philosophy & Economics*, 2008, vol. 7, n° 1, p. 57-80.

Peter VALLENTYNE, « Taking Justice Too Seriously », *Utilitas*, 1995, vol. 7, n° 2, p. 207-216.

Manuel VARGAS, *Building Better Beings: A Theory of Moral Responsibility*, Oxford, Oxford University Press, 2015.

Manuel VARGAS, « Revisionist Accounts of Free Will: Origins, Varieties, and Challenges », in Robert KANE (dir.), *Oxford Handbook on Free Will, 2nd Edition*, Oxford University Press, 2011, p.

Nicholas VROUSALIS, « Exploitation, Vulnerability, and Social Domination », *Philosophy & Public Affairs*, 2013, vol. 41, n° 2, p. 131-157.

Frans B. M. de WAAL, *Peacemaking among primates*, Cambridge, Mass., Harvard Univ. Press, 1989.

Eric-Jan WAGENMAKERS et Quentin GRONAU, « Redefine Statistical Significance Part X: Why the Point-Null Will Never Die ».

Michael WALZER, *Spheres Of Justice: A Defense Of Pluralism And Equality*, New York, Basic Books, 1983.

Simone WEIL, *Œuvres*, Paris, Gallimard, 2000.

Alan WERTHEIMER, *Exploitation*, Princeton University Press, 1999.

S. A. WEST, A. S. GRIFFIN et A. GARDNER, « Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection », *Journal of Evolutionary Biology*, 2007, vol. 20, n° 2, p. 415-432.

B. A. O. WILLIAMS et T. NAGEL, « Moral Luck », *Proceedings of the Aristotelian Society, Supplementary Volumes*, 1976, vol. 50, p. 115-151.

Bernard Arthur Owen WILLIAMS, « Ethical consistency », in *Problems of the self: philosophical papers 1956 - 1972*, Cambridge, Cambridge Univ. Press, 1999, p. 166-186.

Bernard Arthur Owen WILLIAMS, *Problems of the self: philosophical papers 1956 - 1972*, Cambridge, Cambridge Univ. Press, 1999.

Timothy WILLIAMSON, « Philosophical Expertise and the Burden of Proof », *Metaphilosophy*, 2011, vol. 42, n° 3, p. 215-229.

Emily A. WILLOUGHBY, Alan C. LOVE, Matt MCGUE, Jack QUIGLEY et James J. LEE, « Free will, determinism, and intuitive judgments about the heritability of behavior », *Open Science Framework*, 2018.

Jonathan WOLFF, « Fairness, Respect, and the Egalitarian Ethos », *Philosophy & Public Affairs*, 1998, vol. 27, n° 2, p. 97-122.

Robert L. WOOLFOLK, John M. DORIS et John M. DARLEY, « Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility », *Cognition*, 2006, vol. 100, n° 2, p. 283-301.

M. E. YAARI et M. BAR-HILLEL, « On Dividing Justly », *Social Choice and Welfare*, 1984, vol. 1, n° 1, p. 1-24.

Iris Marion YOUNG, *Responsibility for Justice*, New York, NY, Oxford University Press, 2013.

Iris Marion YOUNG, *Justice and the Politics of Difference*, Princeton, NJ, Princeton University Press, 2011.

Michael YOUNG, *The Rise of the Meritocracy 1870-2033*, Penguin Books, 1962.

Alan ZAITCHIK, « On Deserving to Deserve », *Philosophy & Public Affairs*, 1977, vol. 6, n° 4, p. 370-388.

Michael J. ZIMMERMAN, *An Essay on Moral Responsibility*, Totowa, NJ, Rowman & Littlefield Publishers, 1988.

Matt ZWOLINSKI, « Structural exploitation », *Social Philosophy and Policy*, 2012, vol. 29, n° 1, p. 154-179.