

## FAUT-IL MÉRITER DE SOUFFRIR POUR ÊTRE PUNI ? PARFIT ET LA SANCTION PÉNALE

Nicolas Nayfeld

NoSoPhi, Paris 1

Les trois thèmes du mérite, de la souffrance et de la punition – au sens de sanction pénale – occupent une place importante dans le volumineux *On What Matters*<sup>1</sup> (désormais *OWM*) de Parfit, comme l'ont récemment souligné deux articles<sup>2</sup>. Ces thèmes sont reliés de la façon suivante : selon Parfit, *même si* le déterminisme est vrai, *même si* personne ne mérite de souffrir, d'être moins heureux ou d'être puni, la punition<sup>3</sup> est néanmoins justifiée, essentiellement pour des raisons d'utilité, et non pour des raisons de justice (rétributive). Cette longue affirmation résume ce qu'on pourrait appeler l'approche « parfitienne » de la sanction pénale. Le but de mon article est d'examiner de façon critique cette approche de la sanction pénale qui, me semble-t-il, est en partie déficiente.

Je procéderai en trois temps. Dans un premier temps, je considérerai le compatibilisme de Parfit, c'est-à-dire, sa thèse d'après laquelle on peut à la fois être déterministe et dire d'un individu qu'il aurait pu agir autrement. Dans un deuxième temps, je me pencherai sur sa conception du mérite et, notamment, sur deux idées qui semblent lui tenir particulièrement à cœur<sup>4</sup>, à savoir que personne ne mérite de souffrir et que toute souffrance imméritée est intrinsèquement mauvaise. Dans un troisième temps, enfin, j'analyserai l'appendice E du deuxième volume de *OWM*, intitulé « The Fair Warning View ».

### I. Le compatibilisme de Parfit

Pour qu'une punition ne soit pas *inique*, il est essentiel que la personne punie ait pu agir autrement et donc qu'elle ait été, en un sens, libre : ce présupposé semble difficile à remettre

- 
- 1 PARFIT D., *On What Matters*, Oxford University Press, Oxford, 2011 (pour le vol. I et le vol. II), 2017 (pour le vol. III).
  - 2 SMILANSKY S., « Parfit on Free Will, Desert, and the Fairness of Punishment », *The Journal of Ethics*, vol. 20, n° 1-3, 2016, p. 139-148 et ZAIBERT L., « On the Matter of Suffering : Derek Parfit and the Possibility of Deserved Punishment », *Criminal Law and Philosophy*, vol. 11, n° 1, 2017, p. 1-18.
  - 3 Lorsque je parle de la punition (ou de la peine ou de la sanction pénale), il s'agit (sauf précision) toujours de la punition *en général* (c'est-à-dire, de la pratique) et non de la punition d'un individu *en particulier*.
  - 4 « *I believe very strongly, for example, that no one could ever deserve to suffer.* » (*OWM*, II, p. 429, mes italiques.)

en question<sup>5</sup>. Toutefois, est-ce possible si le déterminisme est vrai ? Le déterminisme, en effet, selon Parfit, énonce que « tous les événements sont causalement inévitables, de sorte que, lorsqu'on agit d'une certaine manière, il aurait été causalement impossible d'agir autrement<sup>6</sup> ».

Avant d'exposer son propre compatibilisme, Parfit critique celui de Kant. Kant soutient que les phénomènes, par opposition aux choses en soi (ou noumènes), sont soumis « à la loi de la nécessité naturelle<sup>7</sup> », que « toute action qui se passe dans un point du temps est nécessairement sous la condition de ce qui était dans le temps qui a précédé<sup>8</sup> ». Si nous étions omniscients, nous pourrions prévoir la conduite future d'un homme avec la même certitude que la trajectoire d'un astre<sup>9</sup>. Dès lors, la liberté semble être une chimère, « un concept sans valeur et impossible<sup>10</sup> ». Comment peut-on « encore la sauver<sup>11</sup> » ? La solution kantienne, qu'on peut qualifier de dualiste, consiste à distinguer le sujet comme phénomène, soumis aux conditions du temps, et le sujet comme chose en soi, insoumis aux conditions du temps : en tant que phénomène, je ne suis pas l'auteur de mes actes et ceux-ci ne me sont pas imputables, mais en tant que chose en soi, je suis doué de spontanéité et donc responsable de mes actes. Pour Kant, je peux dire, sans contradiction, que j'aurais pu ne pas faire une action inévitablement nécessaire<sup>12</sup>. Kant donne l'exemple du criminel-né. À première vue, il serait injuste de le punir, car il n'a pas choisi d'être ce qu'il est. Toutefois, selon Kant, il est aussi responsable que n'importe quel homme, car ses méfaits ont « pour fondement une causalité libre<sup>13</sup> » et sont « la conséquence de principes mauvais acceptés librement<sup>14</sup> ».

Parfit reconstruit le raisonnement kantien en sept étapes<sup>15</sup> :

- i. Mon acte ne peut pas être immoral sauf si j'aurais dû agir autrement.
- ii. « Devoir » implique « pouvoir ». J'aurais dû agir autrement seulement si j'aurais pu agir autrement.

---

5 Reid dit même qu'il s'agit d'« un axiome aussi évident que n'importe quel axiome de mathématiques » (REID T., *Essais sur les pouvoirs actifs de l'homme*, trad. G. KERVOAS, E. LE JALLÉ, Vrin, Paris, 2009, p. 260).

6 OWM, I, p. 258.

7 KANT E., *Critique de la raison pratique*, trad. F. PICAUVET, P. U. F., Paris, 1993, 4<sup>e</sup> édition, p. 101.

8 KANT E., *ibid.*, p. 100-101.

9 KANT E., *ibid.*, p. 105.

10 KANT E., *ibid.*, p. 101.

11 KANT E., *ibid.*

12 KANT E., *ibid.*, p. 104.

13 KANT E., *ibid.*, p. 106.

14 KANT E., *ibid.*, p. 106.

15 OWM, I, p. 258-259.

- iii. Mon acte ne peut pas être immoral sauf si j'aurais pu agir autrement.
- iv. Si mon acte était un simple événement spatio-temporel, il serait causalement déterminé et je n'aurais pas pu agir autrement.
- v. Si mon acte était un simple événement spatio-temporel, il ne pourrait pas être immoral.
- vi. Or mon acte est immoral.
- vii. Par conséquent, mon acte n'est pas un simple événement spatio-temporel.

Parfit adresse deux reproches à ce raisonnement. Premièrement, la 6<sup>e</sup> étape est problématique. En effet, comment puis-je savoir que mon acte est immoral et que la morale n'est pas une illusion sans présupposer la conclusion ? Kant pourrait répondre en disant que la conscience de la loi morale est un « fait de la raison<sup>16</sup> », « le fait unique de la raison pure<sup>17</sup> ».

Deuxièmement, il n'est pas nécessaire de postuler, à l'encontre du principe de parcimonie ontologique, que le sujet est à la fois un être sensible et un être intelligible – postulat inintelligible selon Parfit<sup>18</sup> – pour sauver la liberté et, par là, la morale, l'imputabilité, la responsabilité, etc. Il suffit d'analyser correctement l'expression « J'aurais pu agir autrement » pour réaliser que le déterminisme, à la différence du fatalisme, n'exclut pas que cette expression soit vraie.

Parfit distingue deux analyses de l'expression « J'aurais pu agir autrement » : une analyse catégorique et une analyse conditionnelle<sup>19</sup>. D'après l'analyse catégorique, elle signifie qu'il aurait été possible (*it would have been possible*) que j'agisse autrement, même si les conditions avaient été les mêmes. Si c'était réellement le sens de cette expression, nous devrions cesser de l'utiliser, car elle est incompatible avec un principe déterministe – que Parfit tient pour acquis –, à savoir que les mêmes causes produisent les mêmes effets. Fort heureusement, on peut et on doit l'analyser différemment. D'après l'analyse conditionnelle, « J'aurais pu agir autrement » signifie simplement que j'aurais agi autrement, si les conditions avaient été différentes, si je l'avais voulu ou choisi, et « Je n'aurais pas pu agir autrement » signifie que je n'aurais pas agi autrement, même si je l'avais voulu ou choisi. Selon Parfit, cette analyse est compatible avec le déterminisme. On pourrait aller plus loin et dire que cette analyse est non seulement

---

16 KANT E., *Critique de la raison pratique*, op. cit., p. 31.

17 KANT E., *ibid.*

18 OWM, I, p. 269.

19 OWM, I, p. 260.

« compatible » avec le déterminisme, mais elle énonce l'idée même de nécessité conditionnelle : pour qu'un phénomène ait lieu, un ensemble de conditions doit être réuni.

L'analyse conditionnelle de « J'aurais pu agir autrement » est un lieu commun de la tradition empiriste. On la trouve déjà plus ou moins chez Hobbes<sup>20</sup>, Locke<sup>21</sup>, Hume<sup>22</sup>, Mill<sup>23</sup> et beaucoup d'autres. Toutefois, il est également bien connu que cette analyse a été sérieusement fragilisée – pour ne pas dire ruinée – au cours de la deuxième moitié du XXe siècle. Les deux critiques les plus connues – on verra qu'elles se complètent mutuellement – sont celles d'Austin et de Chisholm. Ces deux auteurs prennent pour cible une analyse de Moore, quasiment identique à celle de Parfit.

En effet, Moore soutient que pour qu'une action soit bonne ou mauvaise, il est essentiel que j'aie pu agir autrement. Qu'est-ce à dire ?

« La réponse qui se présente à l'esprit de façon tout à fait évidente, c'est qu'après tout je veux peut-être simplement dire que j'aurais pu le faire, si tel avait été mon choix ; autrement dit, c'est que (pour éviter une possible complication) nous ferions peut-être mieux de dire : “Je l'aurais fait si j'avais choisi de le faire.” En d'autres termes, cela donne à penser que nous utilisons souvent l'expression “j'aurais pu” simplement et seulement comme une abréviation de “je l'aurais fait si j'avais choisi de le faire” »<sup>24</sup>.

Pour Moore, « Il aurait pu agir autrement » ne signifie pas qu'il aurait pu *absolument* ou *de façon absolue* faire autre chose : ce serait contraire au principe de causalité, d'après lequel tout a une cause. Ce qui justifie qu'on traite différemment un individu ayant volontairement commis un délit et un individu ayant involontairement contracté une maladie est que le

---

20 « Celui-là, en effet, est libre de faire une chose à qui il est loisible de la faire s'il en a la volonté, et de s'en abstenir s'il a la volonté de s'abstenir. [...] J'admets cette *liberté* en vertu de laquelle je *peux* si je *veux*... » (HOBBS T., « Au Marquis de Newcastle », in *De la liberté et de la nécessité*, trad. F. LESSAY, Vrin, Paris, 1993, p. 56.)

21 « Quant à la *liberté*, elle est pour l'homme le pouvoir de faire ou de s'abstenir de faire une action particulière, selon que faire ou s'abstenir est effectivement préféré par l'esprit – ce qui revient à dire : selon que l'homme lui-même le *veut*. » (LOCKE J., *Essai sur l'entendement humain*, livre II, chap. 21, §15, trad. J.-M. VIENNE, Vrin, Paris, 2001, p. 384-385.)

22 « Par *liberté*, alors, nous pouvons seulement entendre *un pouvoir d'agir ou de ne pas agir selon les déterminations de la volonté* ; c'est-à-dire, si nous choisissons de rester en repos, nous le pouvons ; si nous choisissons de nous mouvoir, nous le pouvons aussi. Or, cette *liberté conditionnelle* appartient, de l'aveu universel, à tout homme qui n'est pas prisonnier dans les chaînes. » (HUME D., *Enquête sur l'entendement humain*, section VIII, trad. A. LEROY, revue par M. BEYSSADE, GF Flammarion, Paris, 2006, p. 164.)

23 « ce sentiment que nous avons de modifier, *si nous le voulons*, notre propre caractère est celui même de la *liberté morale* dont nous avons conscience. » (MILL J. S., *Système de logique*, trad. L. PEISSE, Pierre Mardaga, Liège, Bruxelles, 1988, p. 425.)

24 MOORE G. E., *Principia Ethica*, trad. M. GOUVERNEUR, revue par R. OGIEN, P. U. F., Paris, 1998, p. 337.

premier aurait pu ne pas commettre son délit, s'il l'avait voulu, tandis que le deuxième n'aurait pas pu ne pas contracter sa maladie, même s'il l'avait voulu<sup>25</sup>.

Dans une célèbre note de « Ifs and Cans », Austin montre, contre Moore, qu'il n'est pas *nécessaire* que « J'aurais fait X, si je l'avais voulu » soit vrai pour que « J'aurais pu faire X » soit vrai. Il prend l'exemple d'un joueur de golf qui rate son putt, s'énerve et dit qu'il aurait pu le rentrer sans problème. Ce n'est pas vrai que s'il l'avait voulu, s'il avait essayé, il l'aurait rentré : *de fait*, il voulait le rentrer, il a essayé de le rentrer, mais il n'a pas réussi. Ce n'est même pas certain que s'il s'était davantage concentré, il l'aurait rentré : *de fait*, il était au maximum de sa concentration. Pourtant, il peut légitimement affirmer qu'il *aurait pu* le rentrer. En effet, il est champion de France, il ne rate quasiment jamais ce genre de putt et il n'a pas fait d'erreur. Par conséquent, ce putt était à *sa portée* ou « *dans ses cordes* », il était *en mesure* de le rentrer. Toutefois, comme l'écrit Austin, « une capacité ou un pouvoir ou une aptitude humaine est intrinsèquement susceptible de ne pas réussir, à l'occasion, et ce sans raison (ou bien la malchance ou la méforme sont-elles parfois des raisons ?)<sup>26</sup> ».

Chisholm, pour sa part, montre, contre Moore, qu'il ne *suffit* pas que « J'aurais fait X, si je l'avais voulu » soit vrai pour que « J'aurais pu faire X » soit vrai<sup>27</sup>. Selon lui, il est possible que « J'aurais fait X, si je l'avais voulu » soit vrai et que, dans le même temps, « J'aurais pu faire X » soit faux. Imaginons – j'adapte son exemple – qu'un militaire, drogué à son insu par un agent secret, tire sur un général. Il est vrai que *si* le militaire l'avait voulu, il aurait agi autrement, c'est-à-dire, n'aurait pas tiré sur le général. Toutefois, il n'est pas vrai qu'il aurait pu ne pas tirer. Pourquoi ? Pour la simple et bonne raison qu'il n'aurait pas pu, sous l'effet de la drogue, *vouloir* ne pas tirer<sup>28</sup>. Pour sortir de cette impasse, on pourrait compléter l'analyse et dire que pour que « Il aurait pu ne pas tirer » soit vrai, il suffit que « S'il l'avait voulu, il n'aurait pas tiré » *et* « Il aurait pu vouloir ne pas tirer » soient vrais. Mais, dans ce cas, on s'engage dans une régression à l'infini, car il faudra appliquer la même analyse à « Il aurait pu vouloir ne pas tirer », puis à « Il aurait pu vouloir vouloir ne pas tirer », et ainsi de suite *ad infinitum*.

---

25 MOORE G. E., *ibid.*, p. 339.

26 AUSTIN J. L., « "Pouvoir" et "si" », in *Écrits philosophiques*, trad. L. AUBERT, A.-L. HACKER, Seuil, Paris, 1994, note 10, p. 189. Ainsi, d'après Austin, quand on affirme que l'accusé aurait pu faire ou ne pas faire X, on fait toujours référence, entre autres, à une capacité générale.

27 CHISHOLM R. M., « La liberté humaine et le moi », in *La responsabilité. Questions philosophiques*, sous la dir. de M. NEUBERG, P. U. F., Paris, 1997, p. 42-43.

28 On pourrait nous objecter que le militaire non drogué n'est pas la même personne que le militaire drogué. Pour répondre à cette objection, il suffirait de changer de scénario : Luther, s'il l'avait voulu, aurait renié ses thèses, mais il n'aurait pas pu, pour des raisons morales, vouloir les renier.

Bref, l'analyse conditionnelle (ou hypothétique) de Moore – qui est aussi celle de Parfit – paraît échouer : pour que « J'aurais pu faire X » soit vrai, il n'est *ni nécessaire ni suffisant* que « J'aurais fait X, si je l'avais voulu, choisi, essayé, etc. » soit vrai<sup>29</sup>. Cela ne signifie pas que la liberté est incompatible avec le déterminisme, que le déterminisme exclut que j'aie pu agir autrement et que la punition soit justifiée. Cela signifie simplement que ni les arguments de Moore, ni ceux de Parfit ne sont concluants.

## II. Souffrance et mérite

Parfit affirme à de nombreuses reprises, dans *OWM*, que personne ne mérite<sup>30</sup> de souffrir. La section 39 s'intitule d'ailleurs « *Why We Cannot Deserve to Suffer* ». Il pense également que personne ne mérite d'être moins heureux<sup>31</sup> ou d'être puni<sup>32</sup>. Bien entendu, Parfit ne nie pas qu'on mérite certaines choses. On peut mériter la gratitude de quelqu'un (si, par exemple, on lui a rendu un service), on peut mériter d'être félicité (si, par exemple, on a bien travaillé à l'école), on peut mériter d'être blâmé (si, par exemple, on s'est mal comporté en public)<sup>33</sup>. En outre, une personne qui agit irrationnellement est une personne qui mérite d'être sévèrement critiquée, d'être qualifiée d'« insensée », de « stupide » ou de « folle »<sup>34</sup>. Ainsi, il y a deux catégories de choses : celles qu'on peut mériter (la gratitude, être félicité, être blâmé, être sévèrement critiqué) et celles qu'on ne peut pas mériter (souffrir, être moins heureux, être puni).

Parfit reconnaît que tout le monde n'accepte pas cette croyance (*belief*). Certains rétributivistes, par exemple, pensent que lorsqu'un malfaiteur souffre, cette souffrance est intrinsèquement bonne ou, du moins, n'est pas intrinsèquement mauvaise<sup>35</sup>. Parfit pense qu'il

---

29 KANE R., *The Significance of Free Will*, Oxford University Press, Oxford, 1998, p. 53.

30 De la même façon qu'on peut distinguer une acception conséquentialiste et une acception déontologique de « approprié », on pourrait distinguer une acception utilitariste (comme celle de Sidgwick, cf. *infra*) et une acception rétributiviste de « mérité ». Parfit ne glisse pas d'une acception à l'autre et par « mérité » il entend toujours « intrinsèquement digne de », comme le montrent les nombreux exemples qu'il donne. En effet, on ne peut pas dire que personne ne mérite (au sens utilitariste du terme) de souffrir et, en même temps, dire que la peine est justifiée d'un point de vue utilitariste. Merci à Laurent Jaffro pour cette remarque.

31 *OWM*, I, p. 264.

32 *OWM*, II, p. 649.

33 *OWM*, I, p. 272.

34 *OWM*, I, p. 36, p. 123.

35 *OWM*, II, p. 429.

existe peut-être un argument capable de les convaincre du contraire, mais qu'on ne l'a pas encore découvert<sup>36</sup>.

Parfit distingue les nécessités logiques (dont la négation est contradictoire, par exemple «  $p \rightarrow p$  ») et les nécessités non logiques. Selon lui, l'affirmation selon laquelle la souffrance imméritée est mauvaise (*undeserved suffering is bad*) et l'affirmation selon laquelle on doit accepter la conclusion d'un argument si ses prémisses sont vraies et s'il est valide, sont des nécessités non logiques : bien qu'on puisse les nier sans contradiction (logique), elles n'auraient pas pu ne pas être vraies (*they could not have failed to be true*). Il est *inconceivable* que la souffrance imméritée ait pu ne pas être intrinsèquement mauvaise : on ne peut pas soutenir qu'« il se trouve que » la souffrance imméritée est mauvaise<sup>37</sup>, car une telle chose est vraie dans tous les mondes possibles<sup>38</sup>. Bref, l'affirmation selon laquelle la souffrance imméritée est mauvaise est une vérité *non contingente*, bien que *logiquement non nécessaire*.

Selon Parfit, ceux qui nient qu'il s'agit d'une mauvaise chose que les gens souffrent lorsqu'ils ne le méritent pas (*it is bad when people suffer in ways that they do not deserve*), ne savent pas ce qu'ils disent. Il donne l'exemple d'un faon pris au piège mourant dans un feu de forêt. Cet être innocent ne méritait pas de souffrir et on ne peut que déplorer cet événement<sup>39</sup>.

Parfit accepte en partie la conception kantienne du souverain bien. Kant demande : « En quoi consiste le souverain Bien ? » Il répond : « Le bien suprême dans l'ordre de la Création est un monde parfait dans lequel tous les êtres raisonnables sont à la fois heureux et dignes du bonheur<sup>40</sup>. » Autrement dit, le souverain Bien réside dans l'union de la vertu et du bonheur. Pour Parfit, il s'agit d'un objectif inatteignable. Toutefois, on peut et on doit, selon lui, essayer de rendre plus heureux les gens qui sont moins heureux qu'ils ne le méritent<sup>41</sup>, par exemple un saint persécuté dans son pays d'origine ou un enfant surdoué atteint d'une maladie congénitale. En revanche, Parfit se dissocie de Kant lorsque celui-ci affirme « qu'un spectateur impartial ne saurait jamais éprouver de satisfaction à voir que tout réussisse perpétuellement à un être que ne relève aucun trait de pure et bonne volonté<sup>42</sup> » et que « la bonne volonté paraît constituer la condition

---

36 OWM, II, p. 569.

37 OWM, II, p. 643.

38 OWM, I, p. 129.

39 OWM, II, p. 566-567.

40 KANT E., *Leçons d'éthique*, trad. L. Langlois, Librairie Générale Française, 1997, p. 77.

41 OWM, p. 245.

42 KANT E., *Fondements de la métaphysique de mœurs*, trad. V. DELBOS, revue par A. PHILONENKO, Vrin, Paris, 1992, p. 56.

indispensable même de la dignité à être heureux<sup>43</sup> ». Le vicieux, pour Kant, ne mérite pas d'être heureux : seul le vertueux mérite de l'être. Le vicieux mérite de souffrir, car le bonheur doit être proportionné à la bonne conduite :

« Mais si quelqu'un, qui se plaît à taquiner et à tourmenter les gens paisibles, tombe mal un jour et est éconduit avec une solide volée de coups, c'est là sans doute pour lui ce que nous appelons *Uebel*, mais chacun y applaudit et considère la chose comme bonne, en soi, quand même il n'en résulterait rien de plus ; bien plus, celui-là même qui reçoit les coups, doit reconnaître dans sa raison qu'*il les a mérités*, parce qu'il voit là rigoureusement mise en pratique, la proportion entre le bien-être et la bonne conduite, que la raison lui présente nécessairement »<sup>44</sup>.

Nous venons d'exposer le point de vue de Parfit, examinons maintenant ses arguments. Il y en a très peu, ce qui peut paraître étonnant venant d'un auteur aussi attaché à l'argumentation.

Son premier argument est que le déterminisme exclut qu'on mérite de souffrir. Pour mériter de souffrir, il faudrait disposer d'une liberté *absolue*, or la liberté compatible avec le déterminisme (et essentielle à la morale et à la responsabilité) n'est qu'une liberté *conditionnelle*<sup>45</sup> – ce qui, on l'a vu, reste à prouver. Si mon action est mauvaise, alors j'aurais pu agir autrement, c'est-à-dire, j'aurais agi autrement, si je l'avais voulu ; mais si je mérite de souffrir, alors je suis la cause première de mes actions, alors je suis l'artisan de mon propre caractère, ce qui est impossible<sup>46</sup>. Parfit s'inspire ici peut-être de Pereboom (qu'il cite en note<sup>47</sup>), puisque celui-ci soutient dans *Living Without Free Will* que le déterminisme exclut qu'on mérite d'être blâmé ou félicité, mais n'exclut pas qu'on qualifie certaines actions de moralement bonnes ou de moralement mauvaises<sup>48</sup>.

Cet argument pose au moins deux problèmes. Premièrement, pourquoi le mérite, qui est une notion *morale*, exigerait-il une *autre* liberté que celle qui est exigée par la morale ? On pourrait très bien soutenir que pour qu'une punition (ou autre chose) soit méritée, il n'est pas essentiel que la personne punie dispose d'une liberté absolue, il faut simplement qu'elle ait pu agir autrement, qu'elle ait eu le *choix*, ce qui – d'après Parfit – n'est pas incompatible avec le déterminisme.

---

43 KANT E., *ibid.*, p. 56.

44 KANT E., *Critique de la raison pratique*, *op. cit.*, p. 63, mes italiques.

45 OWM, I, p. 265 et OWM, III, p. 297.

46 OWM, I, p. 266-267.

47 OWM, I, p. 506.

48 PEREBOOM D., *Living Without Free Will*, Cambridge University Press, Cambridge, 2003, p.141-143.



Deuxièmement, si le déterminisme exclut qu'on mérite de souffrir, alors il exclut également qu'on mérite d'être heureux, d'être blâmé, d'être critiqué, bref, qu'on mérite toutes les choses qu'on *peut* mériter d'après Parfit. Il est incohérent de soutenir à la fois que personne ne mérite de souffrir *si* le déterminisme est vrai et que certaines personnes méritent d'être plus heureuses *même si* le déterminisme est vrai.

Le deuxième argument de Parfit est ce qu'on appelle un *evolutionary debunking argument*, autrement dit, un argument consistant à dévoiler les origines évolutionnaires d'une croyance normative ou évaluative pour la discréditer, pour saper ou affaiblir sa justification<sup>49</sup>. Selon lui, la croyance d'après laquelle il est intrinsèquement bon (ou n'est pas intrinsèquement mauvais) que le criminel souffre sert à justifier – ou n'est qu'une rationalisation de – certaines de nos attitudes réactives naturelles (que nous avons en commun avec certains animaux) comme la colère, le désir de rendre le mal pour le mal, le ressentiment, la malveillance, etc. La théorie de l'évolution peut sans doute expliquer pourquoi nous avons de telles attitudes réactives naturelles et, par là, montrer que ni celles-ci, ni les croyances servant à les justifier ne sont rationnelles<sup>50</sup>.

L'idée que le ressentiment et le désir de rétribution ont un rôle vital est à nouveau un lieu commun. Elle existait bien avant la théorie de l'évolution de Darwin, puisqu'on la trouve déjà plus ou moins chez Butler<sup>51</sup> ou chez Reid<sup>52</sup>. Elle a été défendue dans un article important de Mackie de 1982<sup>53</sup>.

Le problème de ce deuxième argument – et de tous les *evolutionary debunking arguments* en général – est qu'on peut le qualifier de sophisme génétique (*genetic fallacy*). Un sophisme génétique « consiste à tenter de prouver la fausseté d'une proposition sur la base de faits qui portent sur la genèse de la croyance des individus dans la validité de cette proposition<sup>54</sup> ». Ce n'est pas parce qu'une croyance a une origine évolutionnaire – à supposer qu'elle en ait une –

49 KAHANE G., « Evolutionary Debunking Arguments », *Noûs*, 45:1, 2011, p. 104. Parfit écrit : « This may be one of the cases in which an evolutionary explanation helps to undermine what it explains. » (*OWM*, II, p. 429.)

50 *OWM*, II, p. 429.

51 BUTLER J., « Upon Resentment » (Sermon 8), in *Fifteen Sermons Preached at the Rolls Chapel and other writings on ethics*, Oxford University Press, Oxford, 2017, p. 68-75.

52 « Mais tout comme les armes offensives constituent souvent le meilleur moyen de nous défendre en dissuadant l'ennemi de nous attaquer, la nature a fourni à l'homme et aux autres animaux un tel moyen de défense, avec ce ressentiment soudain dont nous sommes en train de parler, principe plus prompt que les déterminations les plus rapides de la raison, principe prêt à s'enflammer en un instant, et qui fait peser sur l'ennemi la menace d'une revanche. » (REID T., *Essais sur les pouvoirs actifs de l'homme*, trad. G. KERVOAS, E. LE JALLÉ, Vrin, Paris, 2009, p. 147-148.)

53 MACKIE J. L., « Morality and the Retributive Emotions », *Criminal Justice Ethics*, 1:1, p. 3-10.

54 GUILLARME B., *Penser la peine*, P. U. F., Paris, 2003, p. 7.

ou sert à justifier une attitude ayant une origine évolutionnaire qu'elle est, de ce fait, fausse ou irrationnelle.

En outre, Parfit n'est pas cohérent, puisqu'il conteste, de façon générale, la pertinence de ce type d'argument. Dans *Reasons and Persons*, il écrit : « Quand une croyance ou une attitude a une explication évolutionnaire, cela, en soi, n'implique rien du tout (*has neutral implications*). Cela ne suffit pas à montrer que cette croyance ou attitude est justifiée ou ne l'est pas<sup>55</sup>. » Cette affirmation est répétée dans le deuxième volume de *On What Matters* :

« Quand l'acceptation de certaines croyances normatives a rendu une communauté ou une culture plus apte à survivre et à prospérer, ce fait, comme tel, ne remet pas en cause la vérité ou la plausibilité de ces croyances. [...] Bien que l'acceptation de ces croyances puisse aider les communautés à survivre et à prospérer, ce fait ne discrédite pas (*does not debunk*) ces croyances »<sup>56</sup>.

Nous venons d'examiner les arguments de Parfit ainsi que leurs failles. On peut adresser plusieurs critiques supplémentaires à son point de vue.

Premièrement, celui-ci ne semble pas être cohérent. Parfit soutient à la fois que personne ne mérite de souffrir ou d'être moins heureux, que la souffrance imméritée est intrinsèquement mauvaise et que certaines personnes sont moins heureuses qu'elles ne le méritent. Or s'il est possible qu'une souffrance soit *imméritée*, alors il doit être possible qu'une souffrance soit *méritée*, contrairement à ce qu'affirme Parfit, sans quoi on ne pourrait plus utiliser le mot « immérité ». De même, s'il est possible que certaines personnes soient moins heureuses qu'elles ne le méritent, alors il doit être possible que certaines personnes soient plus heureuses qu'elles ne le méritent et méritent donc d'être moins heureuses qu'elles ne le sont. Comme le disait Wittgenstein : « La *proposition* positive doit présupposer l'existence de la *proposition* négative, et vice versa<sup>57</sup>. »

Autre incohérence, Parfit soutient à la fois que personne ne mérite de souffrir ou d'être moins heureux et qu'on peut mériter d'être blâmé ou sévèrement critiqué. Or comment peut-on mériter des choses qui impliquent nécessairement de souffrir ou d'être moins heureux, à savoir d'être blâmé ou sévèrement critiqué, si personne ne mérite de souffrir ou d'être moins heureux<sup>58</sup> ? Si personne ne mérite de souffrir ou d'être

---

55 PARFIT D., *Reasons and Persons*, Oxford University Press, Oxford, 1984, p. 186.

56 OWM, II, p. 537.

57 WITTGENSTEIN L., *Tractatus logico-philosophicus*, 5.5151, trad. G.-G. GRANGER, Gallimard, 1993, p. 86.

58 Peut-être qu'en blâmant un individu, on ne cherche pas à le faire souffrir. Peut-être qu'on veut simplement lui rappeler ses devoirs. Toutefois, il serait étrange de dire « Tu ne mérites pas de souffrir, mais tu mérites d'être blâmé » tout en sachant très bien que le blâme causera de la souffrance chez celui qui recevra le

moins heureux, alors les seules choses qu'on peut mériter sont des choses positives, qui n'impliquent ni souffrance, ni diminution du bonheur. On peut éventuellement imaginer une communauté dans laquelle le concept de mérite fonctionnerait de cette façon, dans laquelle les mauvaises notes, les licenciements, les critiques, les reproches, les punitions, etc., seraient toujours considérés comme immérités et susciteraient systématiquement l'indignation, peu importe ce qui a eu lieu. Toutefois, pourquoi faudrait-il vouloir d'une telle tyrannie de la bienveillance ?

Deuxièmement, il ne va pas de soi que l'affirmation d'après laquelle la souffrance imméritée est intrinsèquement mauvaise soit une vérité non contingente, nécessaire et vraie dans tous les mondes possibles (ou plus précisément dans tous les mondes possibles où il y a de la souffrance). Cette affirmation est vraie seulement si la souffrance *peut* être imméritée. Or beaucoup de déterministes soutiennent que, dans un monde intégralement déterminé, la souffrance n'est *ni* méritée, *ni* imméritée, elle n'est qu'un événement causalement inévitable. Parfit lui-même, d'ailleurs, dit que le déterminisme *exclut* qu'on *mérite* de souffrir : pourquoi n'exclurait-il pas qu'on *ne* mérite *pas* de souffrir ?

Dans *The Methods of Ethics*, ouvrage auquel Parfit renvoie dans son chapitre consacré au déterminisme, Sidgwick soutient que si le déterminisme est vrai et si le libre arbitre est une illusion, alors les notions de mérite et de justice doivent être considérablement modifiées<sup>59</sup>. Selon lui, la seule interprétation de la notion de mérite compatible avec le déterminisme est l'interprétation utilitariste d'après laquelle un individu *mérite* une récompense lorsqu'il est *avantageux (expédient)* de le récompenser, c'est-à-dire, lorsque cela *l'incite* à persévérer et incite les autres à l'imiter<sup>60</sup>. Or, d'après cette interprétation, la souffrance *peut* être méritée, car elle est souvent avantageuse.

Certains croyants pourraient même soutenir qu'il n'est pas possible, dans un monde gouverné par Dieu, qu'une souffrance soit imméritée : si un individu souffre, il doit nécessairement avoir fait quelque chose (qu'on ignore peut-être) justifiant qu'il souffre.

Comment expliquer que Parfit ait soutenu un tel point de vue ? Il est possible que Parfit ait cru que si la souffrance *peut* être méritée, alors on ne peut pas échapper au rétributivisme « qui exige qu'on inflige aux criminels la punition qu'ils méritent, même si elle ne bénéficie à personne<sup>61</sup> ». Toutefois, on peut très bien considérer qu'une personne mérite de souffrir (en

---

blâme.

59 SIDGWICK H., *The Methods of Ethics*, Palgrave Macmillan, Londres, 1962, 7<sup>e</sup> édition, p. 284.

60 SIDGWICK H., *ibid.*, p. 284, note 1.

61 OWM, II, p. 216.

raison des atrocités qu'elle a commises) et, en même temps, se refuser à la punir *pour cette raison* (qu'on peut juger insuffisante<sup>62</sup>), s'interdire d'agir sur la base de cette croyance, même si elle est justifiée. Comme l'écrit Hart :

« Même les plus fervents adhérents de la doctrine utilitariste ont dû être tentés par moments par l'idée simple selon laquelle il est bon ou juste qu'un individu ayant intentionnellement fait souffrir autrui soit lui aussi amené à souffrir. Je doute que quiconque, à la lecture des archives d'Auschwitz ou de Buchenwald, n'ait pas ressenti une forte attirance pour ce principe... »<sup>63</sup>

Ce n'est pas parce que les juges estimaient que Eichmann méritait de souffrir – ce qu'ils estimaient certainement selon Hart – qu'ils l'ont exécuté, mais parce qu'il était coupable de crime contre l'humanité.

Le problème est qu'en disant que personne ne mérite de souffrir, Parfit raisonne encore *en termes de mérite*, alors qu'il aurait peut-être dû éliminer cette notion ou ne pas tenir en compte : ou bien parce que son fondement est trop incertain, ou bien parce que le mérite est trop difficile à mesurer. Cette méfiance vis-à-vis de la notion de mérite (ou de démérite) est caractéristique de l'utilitarisme classique, par exemple de Hume<sup>64</sup> ou de Bentham<sup>65</sup> (qui *lui aussi* pensait que toute souffrance est intrinsèquement mauvaise, que toute punition est un mal).

### III. *The Fair Warning View*

Pour Parfit, *même si* toutes nos actions sont de simples événements dans le temps et sont causalement inévitables, *même si* personne ne mérite de souffrir, d'être moins heureux ou

---

62 Comme l'explique J. FEINBERG : « a person's desert of X is always a reason for giving X to him, but not always a conclusive reason... » (« Justice and Personal Desert », in *Doing and Deserving. Essays in the Theory of Responsibility*, Princeton University Press, Princeton, 1970, p. 60.)

63 HART H. L. A., *Law, Liberty and Morality*, Oxford University Press, Oxford, 1968, p. 59. Hart ne parle pas ici de mérite mais le passage se trouve au cœur d'une critique du rétributivisme.

64 « l'incertitude du mérite est si grande, tant à cause de son obscurité naturelle que de la vanité de chacun, que jamais il n'en résulterait une seule règle déterminée de conduite, et la totale dissolution de la société en serait la conséquence nécessaire et immédiate. » (HUME D., *Enquête sur les principes de la morale*, trad. P. BARANGER, P. SALTEL, GF Flammarion, Paris, 1991, p. 96-97.)

65 « C'est le principe d'antipathie qui fait parler de délit comme *méritant* une peine : c'est le principe correspondant de sympathie qui fait parler de telle action comme *méritant* une récompense : ce mot *mérite* ne peut conduire qu'à des passions et à des erreurs. Il ne faut considérer que les effets bons ou mauvais. » (BENTHAM J., *Traité de législation civile et pénale*, Dalloz, Paris, 2010, p. 58.)

d'être puni, la punition – c'est-à-dire, la sanction pénale – est justifiée, à condition qu'elle soit régulière<sup>66</sup> (*fair*).

Parfit distingue, à juste titre, la justice comme *fairness* et la justice comme *desert*<sup>67</sup>. En effet, une punition peut être juste, au sens de méritée, sans être juste, au sens de régulière ; et une victoire peut être juste, au sens de régulière, sans être juste, au sens de méritée. Un exemple de punition *méritée mais irrégulière* serait celle infligée à un criminel nazi en vertu de lois rétroactives (pour certains rétributivistes en tout cas). Un exemple de victoire *régulière mais imméritée* serait celle d'une équipe ayant marqué le plus de points, mais ayant fait preuve d'antijeu durant la totalité du match.

En outre, comme l'a montré Feinberg, il n'y a pas de mérite sans *desert base* : un agent mérite quelque chose *en vertu de faits le concernant* (il a fait de gros efforts, il a été sympathique, il s'est mal comporté, etc.)<sup>68</sup>. À l'inverse, nous n'avons pas besoin de prendre en compte les caractéristiques de la personne subissant un traitement irrégulier pour dire qu'elle en subit un : ce sont les caractéristiques de la *situation* ou des faits objectifs (la loi n'était pas en vigueur au moment des faits, l'équipe a marqué le plus de points, etc.) qui nous autorisent à porter un tel jugement.

Pour Parfit, la punition d'une personne n'ayant pas enfreint la loi est injuste, au sens d'irrégulière et non au sens d'imméritée, et est *presque toujours* immorale. Il laisse entendre que dans certaines circonstances exceptionnelles, elle pourrait ne pas être immorale ou être préférable à la non-punition<sup>69</sup>.

Il ne suffit pas que la répression soit régulière pour qu'elle soit justifiée. Il faut en outre qu'elle soit utile, autrement dit, qu'elle soit suffisamment dissuasive. Si elle ne

---

66 Dans le § 38 de la *Théorie de la justice* consacré à l'État de droit, Rawls parle de « justice comme régularité » (*justice as regularity*). Je pense que c'est ici ce que vise Parfit en parlant de *fairness*. On rappellera qu'en français, régulier peut vouloir dire conforme à la règle ou encore aux attentes de quelqu'un.

67 OWM, II, p. 649.

68 FEINBERG J., « Justice and Personal Desert », in *Doing and Deserving. Essays in the Theory of Responsibility*, Princeton University Press, Princeton, 1970, p. 58.

69 À ce propos, il me semble que la critique que Parfit adresse à Anscombe est déplacée. Anscombe écrit : « si une procédure *consiste* à punir judiciairement un homme pour ce dont on sait clairement qu'il ne l'a pas fait, il ne peut absolument y avoir aucune discussion sur le fait de la décrire comme injuste. [...] Quelqu'un qui essaierait de discuter cela ferait seulement semblant de ne pas savoir ce que signifie "injuste" car c'est un cas paradigmatique d'injustice. » (« La philosophie morale moderne », trad. G. GINVERT et P. DUCRAY, *Klesis*, 9:2008, p. 28.) Selon Parfit, Anscombe dit que la punition d'un innocent est *injuste* au sens de contraire à la justice rétributive et imméritée ; or, d'après Parfit, il ne peut pas y avoir de justice rétributive dans un monde soumis au déterminisme. Cependant, Anscombe ne dit pas que la punition d'un innocent est contraire à la justice rétributive, mais qu'elle est injuste *tout court* et que, celui qui le nierait, ne maîtriserait pas le concept de justice.

dissuadait personne, ou trop peu de gens, elle ferait souffrir en vain, ce qui serait inacceptable.

La conclusion de l'appendice E, « The Fair Warning View », mérite (!) d'être citée dans son intégralité :

« Nous avons souvent plus de raison d'être désolé, non pas pour la victime d'un crime, mais pour le criminel. Comparés à leurs victimes, les criminels ont souvent eu une vie plus défavorisée et misérable. Quand on emprisonne de tels individus, en vue de dissuader du crime, nous devrions regretter vivement ce que nous faisons. Nous devrions porter sur ces criminels le même regard que celui que nous portons sur les gens qui ont été mis en quarantaine parce qu'ils étaient porteurs d'une maladie dangereuse et infectieuse. Le bien-être d'un criminel compte tout autant que le nôtre »<sup>70</sup>.

Si la punition est une pratique à la fois *tragique* et *révoltante*, car elle implique qu'on fasse souffrir des individus qui – selon Parfit – ne méritent pas de souffrir, il est normal qu'on regrette d'être contraint d'y avoir recours pour protéger la société.

L'approche de Parfit de la sanction pénale rappelle évidemment l'approche utilitariste, puisque pour Bentham « toute punition est un dommage, toute punition est en elle-même un mal<sup>71</sup> » et l'exemplarité, c'est-à-dire, la dissuasion, est la fin principale de la punition. Toutefois, elle s'en distingue, dans la mesure où Parfit ne remet pas en cause la punition d'une personne n'ayant pas enfreint la loi pour des raisons *instrumentales*, mais parce qu'elle est injuste au sens d'irrégulière<sup>72</sup>. En effet, il est bien connu que Bentham, dans le chapitre XIII de son *Introduction aux principes de morale et de législation*, soutient que la punition d'une conduite non contraire à la loi est *inappropriée* parce qu'elle ne pourrait être qu'*inefficace*. Hart a bien montré pourquoi l'argument de Bentham ne fonctionnait pas<sup>73</sup> et Parfit a donc raison de condamner la punition de l'innocent, comme Hart, pour des raisons de *justice*.

Toutefois, si l'approche de Parfit évite certaines difficultés de l'utilitarisme, elle en crée de nouvelles. Premièrement, la comparaison que Parfit tente d'établir entre les prisonniers et les

---

70 OWM, II, p. 651.

71 BENTHAM J., *Introduction aux principes de morale et de législation*, trad. du Centre Bentham, Vrin, Paris, 2011, p. 193.

72 Zaibert a peut-être tort d'affirmer : « Regarding punishment, Parfit's views are just variations on Benthamite motifs. » (ZAIBERT L., « On the Matter of Suffering : Derek Parfit and the Possibility of Deserved Punishment », *Criminal Law and Philosophy*, vol. 11, n° 1, 2017, p. 2.)

73 HART H. L. A., « Prolegomenon to the Principles of Punishment », in *Punishment and Responsibility. Essays in the Philosophy of Law*, Oxford University Press, Oxford, 2008, 2<sup>e</sup> édition, p. 19.

personnes mises en quarantaine, rappelant un passage de « Ce que je crois » de Russell<sup>74</sup>, est problématique sur le plan conceptuel et moral. En effet, il existe une différence *conceptuelle* importante entre les *punitions* à proprement parler, d'une part, et les mesures de contrainte ou les mesures de sûreté, d'autre part. La punition implique qu'on porte *délibérément* atteinte à celui qui la subit. Les dommages qu'impliquent les mesures, en revanche, ne sont pas voulus en tant que tels. Bentham exprime clairement cette idée en disant que punir, « *c'est infliger un mal à un individu, avec une intention directe par rapport à ce mal*<sup>75</sup> ». On n'échappera pas à cette difficulté en utilisant le mot « pénalité » à la place du mot « punition », comme le fait Parfit. En outre, en brouillant la frontière entre le *fautif* et le *malade*<sup>76</sup>, Parfit se rapproche dangereusement de ce que Lewis appelait très ironiquement la théorie « humaniste » (*humanitarian*) de la peine<sup>77</sup>, qui a été fortement critiquée pendant la seconde moitié du XX<sup>e</sup> siècle. En effet, en plus de menacer les libertés individuelles, elle dégrade les condamnés, les abaisse au niveau de ceux qui n'ont pas atteint l'âge de raison.

Deuxièmement, si personne ne mérite d'être puni, comme le soutient Parfit, alors on ne devrait punir personne, à moins d'avoir d'excellentes raisons de le faire, raisons qui justifient qu'on s'écarte du mérite ; or il n'est pas certain que le caractère dissuasif de la punition soit une raison suffisante de punir les coupables, dans la mesure où il n'existe pas de preuve empirique claire et non controversée de l'efficacité dissuasive de la sanction pénale.

#### IV. Conclusion

Parfit soutient que la peine est justifiée. Pour qu'elle soit justifiée, il faut que ceux qu'on punit aient pu agir autrement, ce qui, à première vue, est incompatible avec le déterminisme. Parfit pense avoir montré que cette incompatibilité n'était qu'apparente.

---

74 « L'opinion selon laquelle les criminels sont *pervers* et *méritent* d'être châtiés est de celles qu'une moralité rationnelle ne saurait défendre. [...] Faire souffrir un criminel ne saurait se justifier, fût-ce à titre de punition. [...] Je désire simplement suggérer que nous devrions traiter les criminels comme nous traitons ceux qui sont atteints de la peste. Ils constituent les uns et les autres un danger public, et ils doivent être privés de liberté jusqu'à ce qu'ils cessent de constituer un tel danger. Mais l'homme frappé de la peste est un objet de sympathie et de pitié, tandis que le criminel est un objet d'exécration, ce qui est parfaitement illogique. » (RUSSELL B., « Ce que je crois », in *Le mariage et la morale*, trad. G. BEAUROY, G. LE CLECH, Les Belles Lettres, Paris, 2014, p. 276-277.)

75 BENTHAM J., *Théorie des peines et des récompenses*, in *Œuvres de Jérémie Bentham*, édition E. DUMONT, tome second, Société Belge de Librairie, Bruxelles, 1840, 3<sup>e</sup> édition, p. 7.

76 « But the badness of someone's being a cruel murderer is, I believe, relevantly similar to the badness of someone's being insane. » (*OWM*, I, p. 410.)

77 LEWIS C. S., « The Humanitarian Theory of Punishment », *AMCAP Journal*, vol. 13, n° 1, 1987, p. 147-153.

Nous avons vu qu'il ne réussissait pas à prouver la compatibilité du déterminisme et de la liberté.

Parfit ajoute que la peine est justifiée, même si personne ne mérite de souffrir ou d'être moins heureux, même si les criminels ne méritent pas d'être punis. Nous avons vu que les diverses affirmations de Parfit au sujet de la souffrance et du mérite étaient incohérentes et peu argumentées. En disant qu'une souffrance peut être *imméritée*, Parfit ne fait pas barrage au rétributivisme, mais, au contraire, l'alimente.

Parfit pense que seules les peines justes au sens de régulières sont acceptables, même si la peine est justifiée pour des raisons d'utilité. Il distingue la justice rétributive et la justice comme régularité, distinction importante si on veut échapper *et* aux difficultés de l'utilitarisme *et* aux difficultés du rétributivisme.

En définitive, bien que l'approche « parfitienne » de la sanction pénale semble à première vue séduisante, elle résiste difficilement à l'examen. Ce qu'on peut en retenir, en substance, c'est essentiellement une exhortation à la bienveillance, une condamnation de la violence et une aversion pour la souffrance.