

La personne et le tunnel de verre

Sur Derek Parfit, *Reasons and Persons*, et sur l'identité personnelle

Notre vie nous paraît semblable à un tunnel aux parois de verre, parcouru en sens unique. Nous avons, pensons-nous, une connaissance directe de la réalité intérieure de notre tunnel, de notre «monde intérieur». De l'autre côté des parois, il y a le «monde extérieur»: une autre réalité dont nous admettons l'existence, parce que nous la voyons; ou plutôt, parce que nous en voyons une partie et tentons de deviner le reste.

Parmi les choses que nous apercevons au dehors, il y a d'autres tunnels de verre, que nous supposons renfermer d'autres mondes intérieurs, chacun doublement coupé de nous comme nous le sommes d'eux.

Tout cela, nous le supposons, est réel; mais seul nous semble réel au premier degré le monde intérieur de notre tunnel de verre. Par paresse philosophique, peut-être, nous ne sommes pas solipsistes, mais le solipsisme nous paraît bien crédible. Car tout cela dehors, après tout, pourrait bien n'être qu'un rêve...

Le tunnel, là-bas, se termine avec notre mort. Le monde, nous le supposons, existera encore, mais «notre monde», le seul dont l'existence nous paraisse assurée, disparaîtra à jamais. Il y aura une suite, mais *nous* ne la connaissons pas. Il y aura des projets, des peines et des joies, mais ce ne seront pas *les nôtres*; alors, disons-nous à leur propos, «qu'est-ce que cela peut me faire? Pour *moi* ce sera la fin de tout!»

Nous aimerions qu'il n'en soit pas ainsi. Alors nous pensons qu'il y aura «une vie après la mort», c'est-à-dire une vie à *nous* - la seule qui compte vraiment. Ainsi, le mieux que nous parvenions à imaginer est que notre tunnel soit sans fin.

Cette image du tunnel de verre me vient de Derek Parfit. Au début du chapitre 13 de *Reasons and Persons*¹, dans un passage éloquent, il nous dit comment, lorsqu'il croyait en la réalité de l'identité personnelle:

[...] c'était comme si j'étais emprisonné en moi-même. Ma vie me semblait pareille à un tunnel de verre, dans lequel je me serais déplacé toujours plus vite à mesure que passaient les années, et au bout duquel il y avait l'obscurité.

Mais une fois remise en cause cette notion:

[...] les parois de mon tunnel de verre disparurent. Je vis maintenant à l'air libre.

Concernant ses relations avec autrui:

[...] Les autres personnes me sont plus proches. Je suis moins préoccupé par le reste de ma propre vie, et plus par la vie des autres.

Concernant sa mort:

Au lieu de dire: «je serai mort», je devrais dire: «il n'y aura pas d'expériences futures qui seront en relation, de certaines manières, avec ces expériences présentes». En me remettant sous les yeux le contenu réel de ce fait, cette redescription rend celui-ci moins déprimant.

Ces changements profonds dans l'attitude de Parfit envers ce que chacun de nous a de plus intime -

1 PARFIT D., *Reasons and Persons*, Oxford Paperbacks, 1984, p. 281. Les références sans autre indication correspondent à cet ouvrage. Les traductions de passages de cet ouvrage et d'autres ouvrages non publiés en français sont de moi-même.

envers la vie et la mort - résultent simplement de sa perception d'une vérité factuelle. Celle-ci concerne la nature de notre être, et, dirais-je, plus généralement celle des êtres sentients². Parfit a compris qu'il n'existe pas d'identité personnelle. Ou, pour reprendre sa formulation³: que l'identité personnelle qui résulte de nos relations de continuité physique et psychologique se réduit à ces relations; elle n'est pas un *fait additionnel* («*a further fact*»). Parfit se dit ainsi *réductionniste* concernant l'identité personnelle, s'opposant par là au *non-réductionnisme*, que j'appellerai aussi «personnalisme» d'après un texte que je cite plus bas.

L'identité personnelle est l'identité numérique de la personne, ou sujet, au cours du temps. Cet enjeu est en fait celui de l'existence même du sujet: si à chaque fait de sentience correspondait un sujet nouveau, celui-ci se confondrait avec ce fait de sentience et n'aurait aucune existence réelle.

Reasons and Persons est un livre méticuleux dans son raisonnement, d'un style analytique souvent difficile. Malgré cela, certains présupposés ne sont pas toujours explicités, comme l'identification faite d'emblée du choix rationnel à celui concernant les seuls intérêts de l'individu. Certains aussi se fondent simplement sur les intuitions de l'auteur⁴. Si le lecteur ne les partage pas, il peut rester sur sa faim; en particulier, si on ne trouve pas répugnante la «*Repugnant Conclusion*»⁵, laquelle amène Parfit à rejeter l'utilitarisme proprement dit, certains développements de la dernière partie de l'ouvrage, consacrée à la recherche d'une éthique plus impersonnelle, paraissent moins pertinents. D'un autre côté, la question de l'identité personnelle nous touche sans doute de trop près pour que l'on puisse faire abstraction de nos intuitions à son propos. La méthode parfitienne consiste plutôt à prendre celles-ci au sérieux, comme possédant un contenu factuel, pour les confronter à des expériences de pensée, aux réalités de la neurologie et aux exigences de la logique. Ainsi la critique atteint-elle y compris nos évidences intimes, que nous croyions inexprimables autrement que par des tautologies («moi c'est moi»).

Parfit n'est pas le premier à s'être interrogé sur l'identité personnelle. Pour une part, ses conclusions rejoignent celles de Locke ou de Hume auxquels il se réfère. Il nous apprend que l'évidence du «je» cartésien fut contestée par Georg Christoph Lichtenberg:

Lichtenberg affirmait que, dans ce qu'il tenait pour le plus certain, Descartes s'est fourvoyé. Il n'aurait pas dû affirmer qu'un penseur doit être une entité existant séparément. Son célèbre *cogito* ne justifiait pas cette croyance. Il n'aurait pas dû dire: «Je pense, donc je suis». [...] [II] aurait pu affirmer plutôt: «Il est pensé: de la pensée à lieu.» Ou encore: «Ceci est une pensée, par conséquent une pensée au moins est en train d'être pensée⁶.»

Il semble ainsi que sur le sujet de l'identité personnelle il y ait eu aux XVII^e et XVIII^e siècles un important débat aujourd'hui largement oublié. L'utilitarisme, né alors, apparaît comme le lieu naturel de la critique de l'identité personnelle, parce qu'il assigne à chaque agent moral le même but, celui de maximiser le bien collectif, et parce que ce bien est défini sur la base d'une agrégation impersonnelle de faits subjectifs. Pourtant, ce sont aujourd'hui surtout ses adversaires qui reprochent à l'utilitarisme d'être une théorie impersonnelle et de ne considérer les individus que comme des «récipients» à sensations, selon la formulation de Tom Regan⁷, alors même que les utilitaristes semblent avoir largement

2 Les néologismes «sentient» et «sentience» permettent de qualifier les êtres capables de ressentir quelque chose, en évitant les ambiguïtés des mots formés sur «sensibilité» ou «conscience».

3 Par exemple page 209 et suivantes.

4 Voir page 447, où ce fait est discuté.

5 Cette *Repugnant Conclusion* (p. 381 et suiv.) découle de l'impératif de maximisation de l'utilité totale des êtres sentients. À une situation où un nombre donné de personnes jouissent d'une vie de pure extase serait préférable une autre, où seules existent des vies «valant à peine d'être vécue», pourvu que le nombre de ces dernières compense la faible utilité de chacune. Cette conclusion lui semblant inacceptable, Parfit rejette la version totale (*total view*) de l'utilitarisme; pour d'autres raisons, il rejette aussi la version moyenne (*average view*).

6 P. 224. Il s'agit d'une citation partielle de LICHTENBERG C.G., *Schriften und Briefe*, Sudelbuch K (1794?).

7 REGAN T., «The Case for Animal Rights», 1985, traduction «Pour les droits des animaux», *Cahiers antispécistes* n°5 (décembre 1992); article à ne pas confondre avec le livre *The Case for Animal Rights*.

abandonné la remise en cause du personnalisme. Je reviendrai sur le prescriptivisme universel de R.M. Hare, qui se fonde sur un présupposé personnaliste, tout comme ce point de vue exprimé par Henry Sidgwick:

Il serait contraire au sens commun de nier que la distinction entre un individu donné quelconque et un quelconque autre est réelle et fondamentale, et que par conséquent «je» me soucie de la qualité de mon existence en tant qu'individu en un sens, fondamentalement important, en lequel je ne me soucie pas de celle d'autres individus. Ceci étant donné, je ne vois pas comment on pourrait prouver que cette distinction ne doit pas être prise pour fondamentale dans la détermination de la fin ultime de l'action rationnelle d'un individu. [...] [Un homme] peut continuer à soutenir que son propre bonheur est une fin qu'il serait irrationnel de sa part de sacrifier à quelque autre fin que ce soit; et qu'il est par conséquent nécessaire de faire apparaître de quelque manière une harmonie entre la maxime de prudence et celle de bienveillance rationnelle, si l'on veut que la moralité soit entièrement rationnelle. Cette dernière vue [...] est celle à laquelle j'adhère moi-même⁸.

Reasons and Persons traite de sujets importants autres que celui de l'identité personnelle, qui en constitue cependant le noyau et auquel je me limiterai. Le grand mérite de Parfit me semble être d'avoir posé le débat sous l'angle fondamentalement éthique de *ce qui importe*⁹. Car au-delà du discours descriptif sur l'existence d'une entité «je» persistant dans le temps, c'est le plan du prescriptif qui constitue le nœud de la question. Malgré la profonde admiration que Parfit porte à Sidgwick¹⁰, sa position est à l'opposé des thèses énoncées dans le passage ci-dessus. Si Sidgwick a tort et que l'identité personnelle n'est pas «réelle et fondamentale», elle n'est pas ce qui importe, et c'est sur la base d'une croyance fausse que nous nous soucions de la qualité de notre existence d'une manière radicalement différente de celle dont nous nous soucions de celle d'autrui. C'est sur la base d'une croyance fausse que nous craignons tant la mort. C'est sur la base d'une croyance fausse que nous estimons non problématique et rationnellement fondée la prudence, tout en désespérant de trouver à l'éthique un fondement autre que notre propre intérêt bien compris. C'est encore sur la base d'une croyance fausse que nous accordons une place centrale en éthique à des notions personnalistes comme les droits, la justice distributive, la récompense et la punition, etc.

Le terme de «réductionnisme» dont se réclame Parfit concernant l'identité personnelle choque parfois, suggérant une volonté mesquine de nier ce qui est grand. On peut cependant être réductionniste de différentes choses; nous sommes généralement, par exemple, réductionnistes du Père Noël, estimant que celui-ci se réduit aux causes plus banales qui expliquent l'apparition des cadeaux dans les souliers des enfants. Il n'y a donc pas de réductionnisme «en général»; chaque réductionnisme est à juger selon ses propres mérites.

L'emploi du terme par Parfit traduit peut-être une volonté de conserver une distinction entre la thèse de la réduction et celle de l'inexistence; c'est-à-dire, entre l'affirmation que X se réduit à Y, et celle que X n'existe pas, les phénomènes attribués à X étant en fait dus à Y. Cette distinction ne semble cependant fondée que si l'on cherche à préserver, dans la première affirmation, l'idée que X puisse être malgré tout plus que Y, c'est-à-dire, puisse être vu comme «phénomène émergent»; autrement, X n'est qu'un nom abrégé pour le ou les faits Y qui fondent X. Mais Parfit nous dit bien que l'identité personnelle n'est rien de plus que les faits de continuité physique et psychologique. La distinction entre réduction et affirmation d'inexistence est donc peu justifiée; la thèse de Parfit est, en substance, celle de l'inexistence de l'identité personnelle, et s'oppose à la thèse inverse: celle de son existence, en tant qu'entité irréductible aux relations de continuité physique et psychologique.

8 SIDGWICK H., *The Methods of Ethics*, 1874, p. 498, cité en partie par Parfit, p. 138.

9 Voir en particulier ch. 12, «Why Our Identity Is Not What Matters».

10 Cf. l'introduction de PARFIT D., *On What Matters*, ouvrage en préparation, version 4/2008 disponible sur le Web en <http://users.ox.ac.uk/~ball2568/parfit/>.

Il est essentiel de distinguer l'enjeu de l'identité personnelle, ou du sujet, de celui de la sentience, ou de la subjectivité. La thèse de Parfit remet justement en cause l'évidence apparente selon laquelle il ne peut y avoir de subjectivité sans sujet. Or il existe *aussi* un réductionnisme de la sentience, et bien des arguments qui prétendent montrer l'inexistence de celle-ci, par exemple chez Daniel Dennett¹¹, ne valent en fait que contre l'existence du sujet. Le réductionnisme de Parfit, au contraire, loin de viser à nier la sentience, nous permet d'appréhender les événements de sentience comme existant dans le monde, plutôt que comme des entités «privées» n'existant que du point de vue d'un sujet.

Parce que la thèse non-réductionniste se confond avec le sens commun et se trouve rarement énoncée en tant que telle, j'en citerai une expression particulièrement nette:

Au cœur du *personnalisme* du pape Jean Paul II (sa philosophie de la personne) se trouve la reconnaissance du fait que c'est la personne individuelle concrète qui est le sujet de la conscience. [...] Ce sujet peut exister avant la conscience (comme dans le cas de l'embryon humain) ou pendant les interruptions de la conscience (comme au cours du sommeil ou d'un coma). Mais le sujet existant n'est pas à identifier à la conscience elle-même, laquelle est une opération ou activité du sujet¹².

Ce sujet est un objet de nature très spécifique, «mystérieux et unique» selon le *Catéchisme de l'Église Catholique*:

Dans le royaume, le caractère mystérieux et unique de chaque personne marquée du nom de Dieu resplendira en pleine lumière. «Au vainqueur, ... je donnerai un caillou blanc, portant gravé un nom nouveau que nul ne connaît, hormis celui qui le reçoit» (Ap 2, 17)¹³.

Parfit se place au contraire dans une perspective physicaliste, la sentience correspondant à ce qui se passe dans notre cerveau, et l'ouvrage se conclut par un appel au développement d'une éthique athée:

La croyance en Dieu, ou en une pluralité de dieux, a empêché le développement libre du raisonnement moral. La non-croyance en Dieu, admise ouvertement par une majorité, est un événement récent, encore inachevé. Pour cette raison, l'éthique non religieuse n'en est qu'à ses balbutiements¹⁴.

C'est pourtant parmi les rationalistes déclarés que l'on trouve certaines des proclamations personnalistes les plus virulentes, le rationalisme et le matérialisme étant interprétés comme disqualifiant l'éthique et «donc» comme justifiant le «chacun pour soi» sur fond d'exaltation de l'individu et de son autonomie héroïque¹⁵. Sans aller jusque-là, même les irreligieux parmi nous adhérons facilement à la croyance sidgwickienne en un «moi» qui nous sépare, en un sens fondamental, de tout autre «moi». La critique de cette conception m'apparaît comme un des éléments fondateurs d'une éthique non religieuse.

Les relations de continuité

Notre identité personnelle, pourrait-on penser, est simplement l'identité de notre corps au cours du temps. Celle-ci, cependant, n'est pas réellement fondée, et surtout n'est pas vraiment ce que nous avons à l'esprit lorsque nous croyons à l'identité personnelle.

Nous sommes bien identifiés à notre corps dans nos interactions quotidiennes. Cela est

11 DENNETT D., *La Conscience expliquée*, Odile Jacob, 1993. Voir en particulier la critique du «théâtre cartésien» dans le chapitre 5.

12 DEMARCO D., «Peter Singer: Architect of the Culture of Death», http://www.catholiceducation.org/articles/medical_ethics/me0049.html.

13 §2159.

14 P. 454.

15 Un exemple est celui de la romancière Ayn Rand et du courant ultra-libéral «objectiviste» qu'elle a fondé. On pense aussi à l'anarchiste Max Stirner, ou encore aux justifications de l'égoïsme qui se réclament du darwinisme.

opérationnellement possible parce que chaque corps décrit au cours du temps une trajectoire spatiale continue, ce qui permet théoriquement de le suivre à chaque instant et ainsi de lui assigner une identité invariante de manière non ambiguë.

La composition matérielle de notre corps se modifie avec le temps, mais là encore seulement de manière continue. Entre deux instants proches, cette composition ayant peu varié, nous trouvons négligeable l'erreur commise en affirmant qu'il s'agit du même corps matériel. Pourtant, après quelques années notre matière se renouvelle presque entièrement. Peut-on alors dire qu'il s'agit du même corps, au sens matériel?

Cette question a été posée au moins depuis l'Antiquité:

Le vaisseau sur lequel Thésée s'était embarqué [...] était une galère à trente rames, que les Athéniens conservèrent jusqu'au temps de Démétrios de Phalère. Ils en ôtaient les vieilles pièces, à mesure qu'elles se gâtaient, et les remplaçaient par des neuves qu'ils joignaient solidement aux anciennes. Aussi les philosophes [...] soutiennent les uns que c'était toujours le même [vaisseau], les autres que c'était un vaisseau différent¹⁶.

En partant d'un vaisseau A, nous obtenons, par remplacement successif de toutes ses pièces, un vaisseau B dont, s'il avait été construit directement à partir de ces nouvelles pièces, personne n'aurait songé à dire qu'il s'agit du même vaisseau que A. Or la valeur de vérité de la proposition $A = B$ dépend de ce que sont A et B, et non des états intermédiaires. Cette valeur doit donc être la même, que B ait été créé par remplacement progressif des pièces de A ou par assemblage direct des mêmes nouvelles pièces. Puisque dans le second cas, nous ne dirions pas que $A = B$, nous ne pouvons le dire dans le premier. Il en va de même pour notre corps: si après un certain temps toutes ses molécules ont été remplacées, nous ne pouvons dire qu'il s'agit du même corps.

Surtout, la continuité physique ne rend pas compte de la conception réelle que nous avons de notre identité. En effet, que nous croyions ou non au paradis, nous accordons un sens, vrai ou faux mais intelligible, à l'idée d'y aller. Nous croyons que *je* est un objet dont il est peut-être impossible de fait, mais non logiquement, qu'il subsiste sans notre corps. Et ce qui nous importe est cet objet: il nous importe que «nous» allions ou non au paradis, et non que notre corps y aille. Notre identité, cet objet dont l'avenir nous importe tant, et d'une manière si différente de celle dont nous nous soucions de l'avenir d'autrui, n'est pas constitué par notre continuité physique. Celle-ci sert seulement à son repérage quotidien.

Ce «moi» est d'ordre psychologique. Selon John Locke, l'identité personnelle est définie par la mémoire:

Car dès lors qu'un être intelligent peut répéter l'idée d'une quelconque action passée avec la même conscience qu'il en avait à l'origine, et avec la même conscience qu'il a d'une quelconque action présente, dans cette mesure il s'agit du même moi personnel¹⁷.

Parfit ajoute à ce critère d'autres continuités psychologiques, comme la relation entre une intention d'agir et son exécution ultérieure. Mais il note aussi que ces relations doivent être complétées par transitivité: nous nous souvenons rarement d'une journée passée lointaine, mais chaque jour nous nous souvenons (en grande partie) de nos expériences de la veille. La relation est considérée comme établie si elle l'est de proche en proche¹⁸.

Parfit semble accepter que l'identité personnelle puisse valablement se définir sur la base de telles relations de continuité, tout en argumentant que cette identité n'est pas ce qui importe. Je trouve préférable de dire qu'une telle identité ne peut en être une. En effet, les relations de continuité, qu'elles concernent la trajectoire des corps dans l'espace, leur composition matérielle ou nos souvenirs,

16 PLUTARQUE, *Vie de Thésée*, XXIII, trad. Dominique Ricard.

17 LOCKE J., *Essay Concerning Human Understanding* (1690), livre. II, ch. XXVII, sect. 10.

18 P. 205.

souffrent toujours du même défaut: elles font dépendre l'identité $A = B$ non de ce que sont A et B, mais des états intermédiaires entre A et B.

Parce que ces relations de continuité coïncident cependant presque toujours avec l'idée que nous nous faisons de notre identité, il est difficile de mettre en relief ce en quoi nous sommes, de fait, généralement non-réductionnistes. Parfit se tourne pour cela vers des expériences de la pensée où cette coïncidence est brisée.

La téléportation

L'idée de téléportation, telle qu'elle apparaît en science-fiction et est reprise par Parfit¹⁹, implique le transfert non de la matière de notre corps, mais de l'information concernant son état physique instantané. Sur Terre, le voyageur entre dans une cabine et appuie sur un bouton. Un «scanner» enregistre alors la position et l'état de chacune des molécules de son corps, cerveau compris, et envoie cette information par ondes radio à une autre cabine située sur Mars, où elle sert à créer une copie identique à partir de matière disponible sur place. Par ailleurs, le scanner détruit le corps d'origine²⁰. La copie possède la totalité des souvenirs de l'original, exécutera tous ses projets et se perçoit comme la même personne.

Tout se passe comme si, au lieu d'être détruit, le voyageur avait, simplement, voyagé. Pourtant, nous pouvons aussi décrire l'événement comme comportant d'une part la mort de l'original, et d'autre part la création sur Mars d'une copie qui croit, à tort, être l'original. Le critère de continuité physique est rompu, et si la continuité psychologique est maintenue, ce n'est pas par le moyen habituel. Les souvenirs, par exemple, sont transmis de l'original à la copie par une chaîne causale faisant intervenir le scanner, les ondes radio, etc. Si nous acceptons une telle chaîne causale comme pouvant constituer une relation de continuité fondatrice d'identité personnelle, pourquoi n'accepterions-nous pas au même titre une autre chaîne causale, celle qui se produit lorsque nous racontons nos souvenirs à autrui, et qui passe par les ondes sonores, etc.? Où placer la limite? Il semble impossible dans ces conditions de dire si l'identité personnelle, définie par cette continuité, est maintenue ou non.

La question posée par Parfit est si, au moment d'appuyer sur le bouton de départ, le sujet est fondé à se croire sur le point de mourir - et d'entretenir les sombres pensées correspondantes - ou de voyager. Tant de l'extérieur que de l'intérieur, rien ne distingue la téléportation d'un voyage. Je pars le matin sur Mars, et reviens le soir par le même moyen. Je - ou la copie de ma copie - possède un souvenir aussi vif de mon petit déjeuner sur Terre que de l'après-midi sur Mars et du trajet de retour le soir. La copie de ma copie témoignera que tout s'est bien passé; et qui pourrait, ou aurait jamais pu, la contredire? Parfit suggère que la répétition et la banalisation de l'expérience pourraient même nous amener à oublier nos craintes initiales. Aurions-nous raison ou tort?

Nous avons ici une question d'un genre particulier, sur le statut duquel il me semble important de nous arrêter. Il ne s'agit de savoir ni ce qui *est* (question descriptive), ni ce que nous *devons faire* (question prescriptive), mais ce dont nous sommes fondés à nous attrister ou à nous réjouir par avance. La crainte ou le plaisir anticipatoire peuvent, bien sûr, nous servir lorsqu'ils nous poussent à agir; mais la question ici est celle de ces sentiments en l'absence d'une telle utilité. C'est en premier lieu ce domaine, que j'appellerai de sympathie anticipatoire, que recouvre la question parfitienne de ce qui importe: suis-je fondé, au moment d'appuyer sur le bouton du téléporteur, à m'attrister face à ma mort ou à me réjouir par avance du bel après-midi que «je» passerai sur Mars?

19 P. 199 et suiv.

20 Une telle téléportation semble possible dans l'état actuel de nos connaissances théoriques. Par contre, la téléportation conservatrice - où l'original n'est pas détruit - qu'envisage aussi Parfit contredit les lois connues de la mécanique quantique.

Nous pouvons nous demander dans quelle mesure la sympathie anticipatoire est accessible à la raison et aux faits. Nous pourrions dire qu'aucun raisonnement qui conclurait que je dois ou non l'éprouver ne peut la toucher, car elle n'est pas de l'ordre du devoir. Si je serai torturé demain, dois-je en être angoissé? On s'attend à ce que je le sois; mais si tout en étant pleinement conscient de ce qui m'arrivera je conserve ma bonne humeur, tant mieux! Demain je souffrirai, mais pourquoi y ajouter une souffrance présente? Dans quelle mesure peut-on donc demander, comme le fait Parfit, si je suis fondé à me réjouir, ou à m'attrister, en entrant dans la cabine du téléporteur? Ne doit-on pas se borner à constater ce que, de fait, je fais?

Ce serait une erreur, et la question de Parfit est justifiée. Par un aspect, la sympathie anticipatoire n'est pas à l'abri de la raison et des faits. Si je suis malheureux parce que je crois que demain il m'arrivera une chose qui, de fait, ne m'arrivera pas, je le suis à tort. Ce n'est pas qu'il conviendrait d'exhorter mes émotions à changer: c'est que de fait elles changeront. Que ma peur de prendre demain l'avion soit rationnelle ou non, ou d'une rationalité indécidable, elle ne subsistera pas si j'apprends que demain, je ne prendrai pas l'avion. Si elle subsiste alors même que j'ai appris une vérité qui la prive de son fondement factuel, c'est que je ne suis pas parvenu à me convaincre intimement de cette vérité (à m'en «rendre compte»).

La sympathie anticipatoire est donc accessible au cognitif. Celui-ci est à son tour accessible à la raison: ce peut être la raison qui me permet de déterminer les faits.

Si face à une situation je m'attriste, mais face à une autre je me réjouis, il faut qu'entre ces situations il y ait une différence substantielle, c'est-à-dire qui ne relève pas simplement d'une différence de description. Or Parfit montre que dans le cas de la téléportation, comme dans d'autres cas qu'il examine, il est possible de faire des événements une description complète *impersonnelle*; et ceci, y compris dans leurs aspects subjectifs. La téléportation peut se décrire de la manière suivante:

- dans la cabine de départ, il est pensé: «je suis sur Terre et vais appuyer sur le bouton»;
- puis il est pensé, sur Mars: «je suis sur Mars et crois me souvenir d'avoir appuyé sur le bouton».

Une telle description est complète (décrit les faits tant du point de vue matériel que subjectif), et pourtant ne dit rien quant au fait que la personne sur Mars est, ou n'est pas, le voyageur de départ. Il n'y a donc aucune différence substantielle entre les deux descriptions suivantes:

(1) Le voyageur est tué, et un nouvel être sentient, qualitativement mais non numériquement identique, est créé sur Mars.

(2) Le voyageur est transporté sur Mars.

S'il n'y a entre (1) et (2) aucune différence substantielle, c'est que l'identité personnelle ne peut être ce qui importe. Si j'ai intimement intégré les raisonnements qui mènent à cette conclusion, ma sympathie anticipatoire ne pourra s'attrister face à (1) tout en se réjouissant face à (2).

Parfit décrit plusieurs autres expériences de pensée, parfois proches du possible technologique actuel, dans lesquelles une description en termes d'identité personnelle est impossible ou entachée d'arbitraire, alors qu'une description impersonnelle, qui constate les états mentaux présents dans le monde, les localise dans un corps précis, mais ne les attribue pas à un sujet persistant dans le temps, est possible et complète. Citons les expériences de division du cerveau, qui donnent naissance à deux flux de conscience²¹. Parfit imagine que la division puisse être réversible à volonté. La description subjective avant la division - un seul flux de conscience -, pendant le temps de division - deux flux qui

21 P. 245 et suiv. La coupure du corps calleux, qui relie les deux hémisphères, a été pratiquée et semble bien donner naissance à deux flux de sentience.

possèdent chacun les souvenirs du flux unique précédent - et après la réunification - un seul flux, qui possède les souvenirs du flux initial et de chacun des deux flux séparés - n'est pas problématique, alors qu'on serait bien en peine de dire lequel des deux flux, pendant le temps de la division, «est réellement» la personne d'origine.

La dissolution des tunnels

Nous avons vu que les descriptions (1) et (2) ci-dessus sont, en substance, identiques. Voyons ce qui fait que (1) est identique à (2), mais diffère de la proposition suivante:

(3) Le voyageur est tué, et aucune copie n'en est créée.

La réponse est: la création de la copie. Ainsi, si je suis dans la position du voyageur qui s'apprête à appuyer sur le bouton, ce qui va arriver très loin ailleurs me concerne, comme si cela devait arriver à moi; il est même dépourvu de substance de dire que cela arrivera à moi ou non. Il est dépourvu de substance de dire si cela se passera dans mon monde intérieur, ou dans celui d'autrui. En somme: ce qui se passe ailleurs dans le monde me concerne, au même titre que ce qui arrive «à moi». Il n'y a pas un monde intérieur, qui me concerne forcément, et un monde extérieur, qui ne me concerne que si je le veux bien. Mon tunnel de verre n'est qu'illusion. Le solipsisme mou qui marque l'attitude personnaliste n'est pas tenable.

C'est la prise de conscience de ce fait qui a produit les effets décrits par Parfit dans ce passage déjà cité:

Au lieu de dire: «je serai mort», je devrais dire: «il n'y aura pas d'expériences futures qui seront en relation, de certaines manières, avec ces expériences présentes». En me remettant sous les yeux le contenu réel de ce fait, cette redescription rend celui-ci moins déprimant.

Si le «moi» n'existe pas, et que nous en sommes intimement convaincus, il nous est impossible de craindre sa disparition. Nous comprenons qu'il y a bien une vie après la mort: celle des autres, qui est comme la nôtre et tout aussi réelle.

Pourtant, Parfit dit seulement que la mort le déprime *moins*. Il se demande s'il nous est possible de croire réellement au réductionnisme:

Nagel a affirmé que, même si la thèse réductionniste est vraie, psychologiquement il nous reste impossible d'y croire. Je vais donc brièvement passer en revue les arguments que j'ai fournis. Je me demanderai alors si je peux *moi-même* honnêtement dire croire à mes conclusions. Si oui, je supposerai que je ne suis pas le seul [...]²².

Sa réponse est mitigée:

J'aurai beau passer en revue mon argumentation, je crains que je ne parviendrai jamais à dissiper entièrement mes doutes. Au niveau réflexif, ou intellectuel, je resterai convaincu de la vérité de la thèse réductionniste. Mais demeurera toujours en moi à quelque niveau inférieur la tendance à croire qu'il doit nécessairement y avoir une différence réelle entre le fait qu'une certaine personne soit moi, et le fait qu'elle soit quelqu'un d'autre. Une chose semblable se produit lorsque je suis au sommet d'un gratte-ciel. Je sais qu'il n'y a pas de danger. Mais, en regardant en bas de cette hauteur vertigineuse, j'ai peur. C'est une semblable peur irrationnelle que j'aurais si je m'apprêtais à presser le bouton [du téléporteur].

[...]

Il est difficile de croire que l'identité personnelle n'est pas ce qui importe. Si demain une personne doit éprouver une grande souffrance, il est difficile de croire que la question si c'est *moi* qui sera cette personne est sans réponse. Et il est difficile de croire que, si je m'apprête à perdre connaissance, il puisse n'y avoir aucune réponse déterminée à la question «suis-je sur le point de mourir²³?»

22 Sect. 94 (p. 274), «Is the True View Believable?»

23 P. 279 et 280.

Une croyance même partielle suffit cependant à rendre une idée opérante. J'ai beau ne jamais croire tout à fait, en voyant un avion au sol, qu'un tel engin puisse voler, cela ne m'empêche pas d'y monter. Comme Parfit, j'ai «du mal à croire» entièrement à la thèse réductionniste, mais je me sens puissamment aidé par la réflexion sur ce sujet à moins craindre la mort et à me sentir plus proche des autres.

Parfit nous dit que les parois de son tunnel de verre ont disparu et qu'il vit maintenant à l'air libre. La relation exclusive que l'identité personnelle établissait «de moi à moi» - entre instants successifs de mon existence - me coupait des autres et du monde physique lui-même. En réalité, la relation de moi à moi est toujours de même nature que celle de moi à autrui. C'est la causalité de la physique qui assure la continuité structurelle de notre corps et de notre cerveau. C'est cette même causalité physique qui pourrait, dans un téléporteur, assurer la continuité structurelle entre l'original et la copie. C'est encore la causalité physique qui relie nos expériences et idées à celles des autres, lorsque nous leur en parlons. La coupure épistémologique entre l'observateur et le monde extérieur est une illusion: la relation de moi à moi, comme celle de moi à autrui, passe par le monde. Notre propre survie passe par ses lois. Si j'abolissais vraiment toute confiance dans le monde - comme dans le doute cartésien radical - je ne pourrais croire à ma propre existence dans même une seconde.

Quelques conséquences

Les conséquences de la critique de l'identité personnelle sont nombreuses et profondes. J'en esquisserai quelques-unes.

Le rapport de moi-maintenant à moi-futur est de même nature que celui de moi à autrui. L'éthique n'est pas plus problématique dans son fondement que la prudence; ou, dit autrement, la prudence n'est pas *moins* problématique que l'éthique. Nous ne pouvons plus dire que notre sort futur doit logiquement nous concerner simplement parce que c'est *le nôtre*. La prudence est aussi difficile à fonder que l'éthique, et pour les mêmes raisons; il n'est plus possible en fait d'en faire un domaine séparé. Il en résulte une situation d'aporie: je n'ai semble-t-il aucune raison valable de faire quoi que ce soit, tout acte ne pouvant bénéficier qu'à un être futur, qui n'est pas moi-maintenant. Cette aporie est, je pense, féconde, nous forçant à redéfinir notre rapport au monde²⁴.

La prudence et l'éthique se fondent en un domaine unique: celui du prescriptif, ou de l'éthique au sens large. Nous retrouvons là une conséquence de l'utilitarisme classique, hédoniste, qui nous commande de maximiser le plaisir du monde, lequel comprend le nôtre. L'imprudence peut être vue comme moralement condamnable, au même titre que le tort causé à autrui²⁵. On peut à l'inverse conclure à une «démoralisation» de l'éthique: le mal fait à autrui est, comme l'imprudence, une *erreur*, due à au fait que, prisonniers dans notre tunnel de verre, nous ne voyons pas les plaisirs et les peines d'autrui comme pleinement existants.

Si l'utilitarisme classique est impersonnel, les versions plus modernes telles que le prescriptivisme universel de R.M. Hare²⁶ et l'utilitarisme des préférences de Peter Singer²⁷ tendent à réintroduire une large part de personnalisme. Cette tendance me semble intenable, au moins au niveau critique²⁸. J'en esquisserai la raison en prenant l'exemple de Hare. Le prescriptivisme universel exprime bien les devoirs moraux en des termes impersonnels, mais se caractérise par le maintien d'un résidu essentiel

24 Cf. OLIVIER D., «Le subjectif est objectif», *Cahiers antispécistes* n°23, décembre 2003.

25 Cf. p. 318, sect. 106. La formulation de Parfit est plus nuancée.

26 Cf. HARE R.M., *Moral Thinking: Its Levels, Method and Point*, Clarendon Press, Oxford, 1981.

27 SINGER P., *Questions d'éthique pratique*, Paris, Bayard, 1997.

28 Sur la distinction entre niveaux éthiques critique et intuitif, voir HARE R.M., p. 25 et suiv.

d'arbitraire dans la détermination par l'individu de son propre bien:

Certaines préférences, même dans le domaine prudentiel, peuvent bien être plus rationnelles que d'autres [...]; il n'en subsiste pas moins un large minimum irréductible de préférences purement autonomes que la pensée rationnelle ne peut qu'accepter pour ce qu'elles sont, ou seront²⁹.

Toute préférence pour soi-même (*self-regarding*) ne peut concerner que son soi futur; si le rapport de moi-maintenant à moi-futur est de même nature que celui d'autrui-maintenant à moi-futur, le privilège accordé par Hare à moi-maintenant, mais non à autrui-maintenant, dans la détermination arbitraire des préférences concernant moi-futur est injustifiable. Si un résidu d'arbitraire doit subsister, il devrait permettre à chacun de choisir les préférences de chacun; mais nous aurions alors, pour celles-ci, une multiplicité d'arbitres, ce qui est absurde. Cette part d'arbitraire ne peut donc subsister.

Si les propositions éthiques n'ont plus rien d'arbitraire, elles sont susceptibles d'avoir une valeur de vérité. La critique de l'identité personnelle favorise ainsi le réalisme éthique, selon lequel les vérités éthiques font partie, à un certain titre, du monde. Elle nous amène aussi à concevoir la sentience elle-même comme partie du monde. Tant que l'expérience subjective était conçue comme liée par essence à un sujet, elle n'existait que pour celui-ci: elle était «privée». Si autrui prenait en compte mes plaisirs et mes peines, c'était par un acte de foi héroïque, ou par espoir de réciprocité. La critique de l'identité personnelle nous amène à voir que les expériences d'autrui existent à plein titre dans le monde comme les nôtres futures. Ce réalisme de la sentience me semble une condition nécessaire pour résoudre le «problème difficile de la conscience» (selon l'expression de David Chalmers), c'est-à-dire celui de l'existence même de la sentience au sein du monde physique. Au contraire, «mystérieux et unique», le sujet, n'étant perçu ni subjectivement ni objectivement, est hors du monde. Tant que l'existence de la subjectivité est conçue comme impliquant celle du sujet, le «problème difficile» est un problème désespéré.

Nous amenant à prendre au sérieux la sentience, la critique de l'identité personnelle nous porte à prendre en compte *toute* sentience, aussi différente qu'elle soit de la nôtre; et ceci comme si c'était la nôtre, donc à égalité avec la nôtre. Pourtant, la notion d'égalité, telle qu'elle est formulée dans le cadre purement humain, est principalement relative à la personne, dans un sens non réductionniste. On parlera d'égalité de valeur ou de dignité des personnes. La mort apparaît comme un mal absolu, puisqu'elle est la destruction de la personne; l'homicide est le crime suprême. À l'opposé, les animaux non humains n'étant pas vus comme des personnes - ils ne vivent que dans l'instant, dit-on - leur vie ne vaut presque rien, moins par exemple que le plaisir que nous prenons à manger leur corps. La notion d'égalité animale me semble ainsi non formulable dans un cadre personnaliste; la réponse immédiate sera: la vie d'une mouche ne peut valoir celle d'un humain.

L'utilitarisme, qui prône l'égalité non de valeur des personnes, mais de prise en compte de leurs intérêts, permet un net progrès dans ce domaine. Cependant, la version proposée par Peter Singer à la suite de Hare, l'utilitarisme des préférences, réintroduit la notion de personne³⁰, et rend difficile toute prise en compte sérieuse de la grande majorité des êtres sensibles. Il me semble que ce n'est qu'en nous libérant de la notion de personne, par la critique de l'identité personnelle, que nous pourrions aboutir à une vision éthique cohérente, au moins au niveau critique, susceptible de prendre en considération les sentients au-delà des frontières de notre espèce.

Cette critique a de nombreuses autres conséquences politiques et économiques, en particulier relativement à la propriété privée, à la notion de contrat, de mérite, etc. Elle relativise d'autres notions d'identité, comme l'identité nationale. Elle est susceptible d'apporter des bouleversements profonds dans l'ensemble de notre pensée, dans notre vision de nous-mêmes, dans nos relations interhumaines individuelles et collectives, dans nos relations avec les autres sentients et dans la compréhension de

29 HARE R.M., p. 226.

30 Cf. SINGER P., ch. 4.

notre position dans le monde.