

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

DEREK PARFIT ET L'APPROCHE PSYCHOLOGIQUE DE L'IDENTITÉ
PERSONNELLE FACE À LA RÉPLIQUE ANIMALISTE

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN PHILOSOPHIE

PAR
MANOHAR COUSSA

MARS 2019

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.10-2015). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

RÉSUMÉ

Ce mémoire est une analyse de l'évolution de la position de Derek Parfit sur la nature ontologique des êtres humains. Nous identifions trois moments distincts chez l'auteur qui seront étudiés chronologiquement, en soulignant les raisons qui motivent le passage de l'un à l'autre, en particulier les arguments de la position animaliste rivale. Initialement, en continuité avec la tradition lockéenne, Parfit suppose que nous sommes des personnes ; il centre donc son propos sur ce qu'est une personne, sur son critère d'identité ainsi que sur son importance pratique. L'animalisme entre en scène par la suite et met de l'avant des arguments qui questionnent cette assomption : en effet, le lockéanisme implique que les personnes ne seraient pas des animaux, alors qu'il est plus plausible de croire que nous sommes de telles créatures. Le débat s'en trouve transformé et s'oriente alors directement sur notre nature ontologique. Désormais, avant de parler de notre critère d'identité, il faut d'abord déterminer si nous sommes essentiellement des personnes ou des animaux. Parfit développe deux positions en réponse : une première qui concède que nous sommes des animaux et que les personnes ne sont qu'une manière d'y référer ; et une seconde qui revient à un lockéanisme plus classique qui nous identifie à la personne, mais qui en fait une partie de l'animal humain. À travers l'examen de ces positions, nous allons accorder une grande importance à une conception qui est constante chez Parfit malgré les changements : le réductionnisme ontologique de la personne. Selon ce point de vue, l'existence de la personne se réduit à l'ensemble des éléments qui la constituent, de sorte que la personne n'est alors que le *bundle* conceptuel qui les réunit. Après avoir critiqué le réductionnisme de Parfit tout comme l'animalisme classique, nous finirons en proposant une position animaliste qui serait compatible avec bien des apports de l'approche psychologique lockéenne, notamment l'idée que la conscience fait partie intégrante de ce que nous sommes.

Mots-clé : Derek Parfit, identité personnelle, lockéanisme, animalisme, réductionnisme, *bundle theory*, *ego theory*

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

DEREK PARFIT ET L'APPROCHE PSYCHOLOGIQUE DE L'IDENTITÉ
PERSONNELLE FACE À LA RÉPLIQUE ANIMALISTE

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN PHILOSOPHIE

PAR
MANOHAR COUSSA

MARS 2019



TABLE DES MATIÈRES

RÉSUMÉ.....	v
INTRODUCTION.....	1
CHAPITRE I	
LA PERSONNE COMME <i>BUNDLE</i> D'EXPÉRIENCES.....	11
1.1 – <i>Parfit I</i> : le réductionnisme constitutif.....	12
1.1.1 – Réductionnisme de l'existence.....	12
1.1.2 – Réductionnismes de l'identité et de son importance.....	18
1.1.3 – Portée du triple réductionnisme.....	22
1.1.4 – L'animalisme comme alternative au réductionnisme.....	24
1.2 – La <i>ego theory</i> peut-elle être défendue ?.....	26
1.2.1 – Le dualisme cartésien, seule version plausible de la <i>ego theory</i> ?.....	27
1.2.2 – L'animal et le réductionnisme constitutif.....	33
1.2.3 – L'animalisme et nos croyances à propos de notre identité.....	38
1.3 – La division et l'unité de la conscience.....	46
1.3.1 – L'argument de la division des hémisphères cérébraux.....	48
1.3.2 – L'argument de la division contesté.....	54
1.4 – Le réductionnisme constitutif et la réification des expériences.....	60
CHAPITRE II	
L'ÉVOLUTION DE LA POSITION DE PARFIT.....	71
2.1 – La réplique animaliste.....	71
2.1.1 – L'argument du fœtus et l'importance de notre vie mentale.....	72
2.1.2 – L'argument de l'animal pensant et le constitutionnalisme.....	76
2.1.3 – La réponse animaliste : la personne comme <i>sortal de phase</i>	83
2.2 – <i>Parfit II</i> : la récupération lockéenne des principes animalistes.....	89
2.2.1 – Le rapprochement animaliste de Parfit.....	90
2.2.2 – Les problèmes de l'animalisme classique.....	95
2.2.3 – Le cerveau comme organisme animal.....	98
2.2.4 – Les critères possibles de notre identité (animale).....	104

2.3 – <i>Parfit III</i> : la personne lockéenne comme partie de l'animal.....	108
2.3.1 – L'animal non-pensant de Shoemaker.....	110
2.3.2 – Le problème des parties pensantes.....	113
2.3.3 – La personne comme partie pensante incarnée.....	117
CHAPITRE III	
UNE ALTERNATIVE ANIMALISTE.....	123
3.1 – Critique de la position finale de Parfit.....	123
3.1.1 – Le réductionnisme de la personne et le penseur non-dérivatif.....	124
3.1.2 – Le problème des parties agissantes.....	129
3.2 – Un argument pour un animalisme compatibiliste.....	137
3.2.1 – L'animalisme compatibiliste.....	138
3.2.2 – Le cerveau et la conscience phénoménale.....	146
3.2.3 – La conscience et le critère d'identité biologique.....	154
3.2.4 – La conscience, les contenus mentaux, et ce qui importe.....	160
CONCLUSION.....	169
ANNEXE A.....	177
BIBLIOGRAPHIE.....	179

RÉSUMÉ

Ce mémoire est une analyse de l'évolution de la position de Derek Parfit sur la nature ontologique des êtres humains. Nous identifions trois moments distincts chez l'auteur qui seront étudiés chronologiquement, en soulignant les raisons qui motivent le passage de l'un à l'autre, en particulier les arguments de la position animaliste rivale. Initialement, en continuité avec la tradition lockéenne, Parfit suppose que nous sommes des personnes ; il centre donc son propos sur ce qu'est une personne, sur son critère d'identité ainsi que sur son importance pratique. L'animalisme entre en scène par la suite et met de l'avant des arguments qui questionnent cette assumption : en effet, le lockéanisme implique que les personnes ne seraient pas des animaux, alors qu'il est plus plausible de croire que nous sommes de telles créatures. Le débat s'en trouve transformé et s'oriente alors directement sur notre nature ontologique. Désormais, avant de parler de notre critère d'identité, il faut d'abord déterminer si nous sommes essentiellement des personnes ou des animaux. Parfit développe deux positions en réponse : une première qui concède que nous sommes des animaux et que les personnes ne sont qu'une manière d'y référer ; et une seconde qui revient à un lockéanisme plus classique qui nous identifie à la personne, mais qui en fait une partie de l'animal humain. À travers l'examen de ces positions, nous allons accorder une grande importance à une conception qui est constante chez Parfit malgré les changements : le réductionnisme ontologique de la personne. Selon ce point de vue, l'existence de la personne se réduit à l'ensemble des éléments qui la constituent, de sorte que la personne n'est alors que le *bundle* conceptuel qui les réunit. Après avoir critiqué le réductionnisme de Parfit tout comme l'animalisme classique, nous finirons en proposant une position animaliste qui serait compatible avec bien des apports de l'approche psychologique lockéenne, notamment l'idée que la conscience fait partie intégrante de ce que nous sommes.

Mots-clé : Derek Parfit, identité personnelle, lockéanisme, animalisme, réductionnisme, *bundle theory*, *ego theory*

INTRODUCTION

Que sommes-nous ? De toute évidence, nous sommes des êtres humains. Mais qu'est-ce que ça veut dire exactement ? Le concept apparaît ambigu lorsqu'on l'examine, rendant le référent du « je » incertain. Les sciences naturelles nous proposent une compréhension de l'être humain dans son appartenance à l'ordre naturel, comme *animal humain*. Cette conception de nous-mêmes ne date cependant pas d'hier, Aristote la soutenait en définissant l'être humain comme animal rationnel. Chacun peut d'ailleurs directement constater combien notre nature et nos capacités sont en continuité avec le reste du monde animal, il suffit d'avoir un animal de compagnie : comme eux, nous sommes des organismes ayant des besoins de subsistance, des désirs, des intentions d'agir pour les combler, et une existence rattachée au bon fonctionnement de nos organes, pour le moins. Le développement de la biologie, de la médecine, ainsi que de la théorie de l'évolution de Darwin sont autant d'appuis à la classification d'Aristote.

Notre mode de fonctionnement en société suggère une saisie de nous-mêmes différente, comme *personnes humaines*, insistant plutôt sur ce qui nous distingue des (autres) animaux. Les facultés intellectuelles et la conscience de soi plus développées chez nous nous permettent de dépasser le comportement purement instinctif : notre rationalité nous rend capables de suivre des règles d'agir socialement établies, normatives ; et notre capacité de réflexion sur notre agir et son impact fait en sorte que nous pouvons chacun en être tenus responsables et être jugés pour avoir respecté ou non les règles d'agir et de moralité. L'être humain est jugé sur ses actes, de sorte que l'historique de son agir, ses dispositions et manières habituelles de se comporter, constituent son caractère et sa personnalité.

La compréhension de l'être humain comme animal et comme personne n'épuise toutefois pas notre conception de nous-mêmes. Notre vie pratique, quotidienne, témoigne notamment de notre nature comme *agents*, c'est-à-dire comme des individus qui produisent des effets dans le monde par impulsion interne, étant eux-mêmes source de leurs actions plutôt que contrôlés à distance tels des pantins. Notre activité mentale

atteste de notre nature comme *penseurs*, comme des êtres ayant une dimension psychologique, pour lesquels des événements mentaux ont lieu, et qui ont de façon générale la faculté de se représenter des états de choses, d'avoir des intentions, des croyances et des sensations, entre autres. Finalement, l'être humain peut être conçu sous sa dimension affective, en tant que *sujet d'expérience*. Nous ne sommes pas seulement des êtres pour lesquels des processus et des événements physiques et mentaux ont lieu, mais nous sommes affectés positivement ou négativement par ceux-ci. Lorsque nous vivons une expérience, nous en avons un *feeling* subjectif, nous sommes capables de ressentir du plaisir et de la douleur, voire d'éprouver des affections plus complexes.

Il y a bien évidemment des recouvrements entre ces divers concepts qui soulignent chacun un aspect de notre conception vulgaire de l'être humain. À première vue, il semble difficile d'établir si les derniers concepts (agent, penseur et sujet d'expérience) se subsument chacun sous la catégorie d'animal ou de personne, sous les deux, ou encore s'ils sont plus larges que ceux-ci. Certains de ces concepts ont été historiquement associés ou dissociés selon notre compréhension changeante de la nature et de nous-mêmes. Descartes, par exemple, défendait une conception mécaniste de la nature qui associait tous ces concepts à celui de personne, ne reconnaissant pas la dimension psychologique des animaux ni leur agentivité – quoique ce soit débatable en ce qui concerne leur affectivité (Harrison 1992). La question qui nous préoccupe dans ce mémoire est la suivante : ces diverses manières de saisir ce que nous sommes sont-elles toutes compatibles ? En particulier, pouvons-nous affirmer de manière cohérente que nous sommes à la fois des animaux et des personnes, ou nous faut-il choisir ?

Ce questionnement se situe dans la problématique philosophique de l'identité personnelle. Il est avant tout à distinguer des questions d'identité ordinaires, qui visent la question du « *qui suis-je ?* », à laquelle on répond par divers attributs non-essentiels, qui ne sont pas nécessairement vrais durant toute notre existence (parent, québécois, colérique, membre de Costco, etc.), plutôt que celle du « *que suis-je ?* » qui nous intéresse ici. Cependant, cette dernière question ne s'est pas toujours trouvée au premier plan du débat philosophique, qui s'est traditionnellement construit autour de la question de l'identité dans le temps, du « qu'est-ce qui fait que je suis le même individu à deux moments différents de mon existence ? ». L'enjeu principal était de déterminer les critères

de notre identité, ou, pour être plus clair, nos *conditions de persistance*.¹ Les diverses réponses à la question de notre *nature ontologique* n'y figuraient que comme conséquences découlant du questionnement central sur ces conditions. Ce n'est que récemment que le débat s'est déplacé pour porter directement sur la question de notre nature ; nous allons voir pourquoi.

L'auteur le plus influent dans le débat traditionnel est sans nul doute Locke. Sydney Shoemaker affirme même que, tout comme A.N. Whitehead a avancé que l'entièreté de la philosophie occidentale n'est qu'un ensemble de notes de bas de page à Platon, il en va de même pour le sujet de l'identité personnelle et Locke (Shoemaker 2008 : 314). Locke a établi l'approche psychologique de l'identité personnelle comme position dominante. Son argument peut être résumé comme suit. Notre concept général d'identité nous provient de l'activité de comparaison d'une chose à différents temps, à laquelle on attribue les mêmes idées, et qu'on considère alors comme la même chose (Locke 1975 : II.XXVII, §1). C'est donc une opération mentale, et celle-ci requiert qu'on ait la faculté de mémoire afin de pouvoir procéder à une telle comparaison. Par suite, on ne peut personnellement s'assurer de notre propre identité qu'à l'aide de notre mémoire. C'est cette mémoire ou conscience de nos actions présentes et passées qui fait de nous nous-mêmes, qui fait qu'on considère ces actions comme les nôtres, nous en rend intimement responsables, et fonde alors notre souci de soi. Notre critère d'identité, c'est donc l'identité de notre mémoire selon Locke, garante de l'identité de notre conscience.

Or, puisque Locke définit la personne comme « un être pensant et intelligent, capable de raison et de réflexion, et qui peut se considérer soi-même comme le même, comme une même chose qui pense en différents temps et en différents lieux » (Locke 1822 : II.XXVII, §9), alors ce que nous réidentifions à l'aide de notre mémoire et que nous sommes essentiellement, c'est la personne. Cette association entre le soi et la personne est ainsi historiquement cimentée, sans équivoque : « *Person*, as I take it, is the name for this *self*.² » (Locke 1975 : II.XXVII, §26) La personne est le nom public donné

1 Parler de critère d'identité peut porter à confondre les *preuves* de l'identité avec ce en quoi elle consiste – ce que Leibniz reprochait à Locke. Selon Leibniz, le critère d'identité personnelle proposé par Locke, l'identité de la mémoire, n'était qu'un critère d'*identification*, c'est-à-dire une manière pour la personne de s'assurer de son identité, sans toutefois être son principe d'individuation ontologique. (Leibniz 1966 : II.XXVII, §1 et §9)

2 La traduction française de Coste comporte une altération majeure au sens original. Je cite le passage traduit : « Je regarde le mot de personne comme ayant été employé pour désigner

au soi dont nous sommes chacun intérieurement assurés dans notre conscience. Ainsi, le critère d'identité (ou plutôt d'identification, voir la note 1) a donc déterminé notre nature ontologique : nous sommes dès lors conçus comme des entités définies par leur psychologie, par leur conscience, c'est-à-dire des *personnes*.

Mais que sont exactement les personnes ? Des *penseurs* conscients de leur identité, nous dit Locke, mais quel type d'entité sont-elles ? La définition lockéenne est bien maigre, c'est à peine si l'on peut dire qu'il leur accorde une nature ontologique quelconque. Est-il possible de préciser davantage cette nature ? Locke va répondre par la négative. Nous, les personnes, ne sommes *que* notre conscience. Sa méthode pour le montrer consiste à tester nos intuitions sur ce que sont les personnes à l'aide d'expériences de pensée. Il passe en revue toutes les possibilités, toutes les entités auxquelles la personne pourrait être associée, et trouve des objections pour chacune. Résumons. L'identité de la personne ne dépend pas de la matière qui la compose, de la substance matérielle, parce que nous sommes encore nous-mêmes si l'on perd un membre, par exemple. Elle ne dépend pas non plus de l'identité de l'âme, de la substance immatérielle – si une telle chose existe. On peut en effet imaginer qu'une personne puisse persister malgré qu'elle change d'âme, sa conscience étant transmise d'âme en âme – comme elle persiste aux changements de matière qui la compose. Associer l'identité de la personne à celle de l'animal, que Locke conçoit comme un corps organisé de façon à soutenir la vie (Locke 1975 : II.XXVII, §4-6), n'a pas plus de succès : il est conceptuellement possible que la personne survive à la mort de l'animal, par résurrection par exemple. Finalement, l'identité personnelle n'est pas celle de l'animal humain, qu'il définit comme la combinaison d'un corps animal et d'une âme, pour les mêmes raisons qu'elle n'est pas ni celle de l'animal ni celle de l'âme pris séparément. On peut de plus imaginer un humain habité par plus d'une conscience distincte – Locke parle des consciences différentes lors du sommeil et de l'éveil, de l'ivresse et de la sobriété, entre autres, où la mémoire n'est pas (pleinement) partagée ; un exemple plus moderne serait celui des personnalités multiples. Bref, conclut Locke, puisque nos conditions de

précisément ce qu'on entend par le mot *même*. » (Locke 1822 : II.XXVII, §26) L'erreur réside ici dans la traduction de « self » par « même », alors que le sens dicterait plutôt « soi-même » ou simplement « soi ». Il est fort probable que le terme ait été d'abord traduit par « soi-même », puis qu'on ait supprimé « soi- » plutôt que « -même », l'erreur ayant ensuite été reproduite dans toutes les éditions françaises subséquentes.

persistance ne semblent pas consister en celles de substances ou d'animaux, l'identité personnelle ne réside que dans l'identité de conscience.

Locke est cependant de l'opinion que la personne est rattachée à *une* âme, et que le scénario où la conscience passe d'âme en âme n'est qu'imaginaire (Locke 1975 : II.XXVII, §25). La relation de la personne à une telle hypothétique substance est toutefois conceptuellement impossible à déterminer faute de preuve empirique sur la nature des âmes – si une telle preuve est même possible – de sorte que le concept de personne en demeure indépendant. Thomas Nagel reconnaît pareillement que le concept de soi ou de personne paraît logiquement distinct de toute substance, mais juge cette « pureté » suspicieuse (Nagel 1986 : 32). L'indépendance conceptuelle du soi ne veut pas dire qu'il n'a pas de référent objectif, c'est-à-dire un objet dans le monde auquel il réfère et qu'il reste encore à trouver : « I suggest that the concept of the self is open to objective "completion" » (Nagel 1986 : 40). Contrairement à Locke, le référent qu'il propose est le cerveau plutôt que l'âme. Mais comme cet hypothétique référent continue de nous éluder, l'approche psychologique de Locke nous conduit donc à la conclusion que le concept de *personne humaine* ne réfère pas à celui d'*animal humain*. Nous, personnes, ne sommes pas des animaux, pas des organismes vivants, parce que nos conditions de persistance pourraient ne pas coïncider avec les leurs. La classification de nous-mêmes dans l'espèce *homo sapiens*, depuis Carl von Linné et tributaire d'Aristote, serait fautive.

Derek Parfit, le philosophe sur lequel on va se concentrer dans ce mémoire, est considéré, avec Shoemaker, comme un des principaux héritiers modernes de Locke. Comme lui, il adopte une approche psychologique de l'identité personnelle, soutient que nos conditions de persistance sont psychologiques et est reconnu pour ses expériences de pensée modernisées qui testent nos intuitions sur ce qu'on est. Ces expériences tendent tout autant à montrer que la personne est conceptuellement distincte de son support matériel. L'idée d'une âme, d'un support immatériel, est cependant rejetée, remplacée par son équivalent insubstantiel moderne : un esprit numérisable, téléchargeable dans tout corps compatible. Pourtant, malgré la réception des écrits de Parfit, son intention n'a pas été de défendre une position lockéenne classique qui dissocie la personne de l'animal :

« In my earlier writings about these questions, my main aim was not to defend a Lockean criterion of personal identity, but to argue that, in our thoughts about our identity, or what is involved in our continuing to exist,

most of us have, at some level, various false beliefs. » (Parfit 2012 : 26)

Plutôt, il a cherché à montrer, contre Locke, que notre conception de l'identité personnelle comme entièrement présente ou absente était erronée, tout comme notre intuition que c'est notre identité qui importe et qui fonde le souci de soi et l'agir rationnel et moral. C'est seulement dans son dernier article sur le sujet qu'il en vient à défendre explicitement que nous ne sommes pas des animaux, poussé dans cette direction par les objections des animalistes qui ont dévoilé la grande difficulté voire l'impossibilité à rendre sa position compatible avec celle qu'on serait des animaux humains.

L'animalisme, malgré la taxinomie introduite par Aristote et les développements subséquents, est une position toute récente, son nom lui ayant été donné par Paul Snowdon en 1991. Auparavant, il était communément admis que ce que nous sommes essentiellement, ce sont des personnes, faisant alors de toute question sur notre nature une question sur ce qu'est une personne, et donc de tout problème sur notre identité fondamentale un problème d'identité *personnelle*. L'animal ne se trouvait nulle part dans le débat ; la question était close depuis Locke. Certes, il y avait une opposition à l'approche psychologique, bien que marginale, mais toujours dans le cadre de la problématique de l'identité de la personne dans le temps. Bernard Williams, notamment, défendait un critère d'identité physique des personnes, plutôt que psychologique comme on le fait depuis Locke – plus précisément, l'identité personnelle résiderait dans l'identité du corps selon lui (Williams 1956). C'était cependant une bataille perdue d'avance : l'idée cartésienne que nous sommes d'abord et avant tout des *penseurs*, et donc que nos conditions de persistance doivent être psychologiques plutôt que physiques ou biologiques, demeure profondément ancrée dans notre façon de nous concevoir, simplement en raison de la séparabilité conceptuelle de notre esprit, et donc de la personne, d'avec le corps et toute substance imaginable. Cela, ajouté à la force et l'exhaustivité des expériences de pensée de Parfit dans *Reasons and Persons* (dorénavant *R&P*), tout comme celles des arguments fonctionnalistes de Shoemaker à propos de l'esprit, a contraint l'opposition de se réinventer avec l'animalisme. Cette position se promet d'aller au fond du désaccord avec le néo-lockéanisme, prenant appui sur les conclusions contre-intuitives de ce dernier.

Le problème principal de l'approche psychologique, lorsqu'on la combine au rejet

naturaliste de l'existence des âmes, c'est la conséquence que nous, personnes, ne sommes pas des choses dans le monde. Le néo-lockéanisme nous définit comme des entités purement psychologiques, dont les conditions de persistance sont purement psychologiques, tout en niant qu'il existe de telles choses dans le monde. C'est un dualisme sans la substance immatérielle. Si l'on faisait l'inventaire de tout ce qui a une existence propre, *nous* n'y figurerions pas, et ce pas même à titre d'objets matériels. De notre nature ontologique, il ne reste donc que le concept, qu'un fantôme sans la machine. Cela pose particulièrement problème à tous ces auteurs qui se targuaient de réduire la personne, l'illusion de l'égo, au nom du matérialisme. Cette conception de nous-mêmes apparaît incompatible avec notre conception du monde guidée par la science, physicaliste, selon laquelle les événements mentaux sont dépendants ou surviennent sur des événements physiques, et où les propriétés psychologiques sont physiquement réalisées. Elle semble également irréconciliable avec le fait que les (autres) animaux, étant eux des objets physiques dans le monde, ont également des propriétés mentales, une vie psychologique – même si plus rudimentaire que la nôtre – tout en ayant des conditions de persistance biologiques plutôt que psychologiques. Serait-il possible que nous soyons fondamentalement des animaux plutôt que des personnes, et qu'alors nos conditions de persistance soient biologiques ?

C'est ainsi que la force longtemps incontestable de l'approche psychologique fut sapée, et que l'animalisme s'introduisit dans l'arène comme opposant principal au néo-lockéanisme. Dorénavant, avant de s'avancer sur les critères d'identité ou les conditions de persistance, il faut résoudre la question de ce qu'on est fondamentalement, sans quoi les faux pas sont inévitables. Affirmer que nous sommes essentiellement des personnes, aux conditions de persistance purement psychologiques, c'est désormais dire que nous n'avons jamais été des embryons, et que nous ne pourrions jamais tomber dans un état végétatif. La question de notre *nature ontologique* a pris le dessus sur celle des *conditions de persistance*, renversant le débat traditionnel sur l'identité « personnelle ».

La réponse des tenants de l'approche psychologique ne se fait pas attendre. On se précipite pour rendre le néo-lockéanisme compatible avec l'idée que nous sommes des animaux. Shoemaker continue de défendre que nous sommes essentiellement des personnes aux conditions de persistance purement psychologiques, mais joue sur le sens

du verbe être pour affirmer que les personnes « sont » des animaux – au sens de « constitués par ». Parfit commence par inclure le cerveau dans nos conditions de persistance, puis adopte la position que nous sommes essentiellement des animaux et s'attaque même à la position « puriste » de Shoemaker (Parfit 1999). Être une personne signifierait alors être un animal ayant des capacités mentales spécifiques. Ce n'est là qu'un aperçu des premières réactions à la critique animaliste, relançant le débat.

L'objectif de ce mémoire est multiple : nous allons nous concentrer sur les écrits de Parfit, afin de voir la transformation du débat annoncée plus haut, à travers les changements itératifs qu'il apporte à sa position. Le premier chapitre va se concentrer sur la position initiale de Parfit présentée en 1971 et développée dans *Reasons and Persons*, avant l'entrée en scène des objections animalistes – ou du moins, avant leur popularisation et que Parfit en prenne connaissance. Notamment, nous allons accorder une grande importance au réductionnisme *ontologique* que Parfit défend et auquel même Shoemaker s'oppose. Nous allons également voir certaines des critiques à sa position qui sont pertinentes pour le débat qui nous intéresse. Ensuite, dans le second chapitre, nous allons nous concentrer sur les critiques animalistes qui poussent Parfit à altérer sa position. Nous présenterons ce faisant la position d'Eric Olson, l'opposant animaliste principal de Parfit, avant d'examiner les positions qu'adopte Parfit en réaction. Ce dernier est d'abord conduit à se rapprocher de l'animalisme en affirmant que la personne n'est qu'un qualificatif de l'animal humain, avant de rebrousser chemin et d'assumer pleinement que nous ne sommes pas des animaux, mais plutôt que les personnes que nous sommes sont une *partie* de ces animaux.

Au terme de l'examen des arguments de part et d'autre, nous allons proposer dans le troisième chapitre une position animaliste alternative qui tente d'éviter les difficultés majeures de chacun des camps, et qui nous permettrait d'affirmer de façon cohérente que nous sommes à la fois des animaux et des personnes, mais qui vient bien sûr avec son propre lot d'obstacles. Le problème principal du néo-lockéanisme, c'est qu'il implique que nous ne sommes pas des êtres matériels, ou, du moins – si la position est amendée pour inclure un référent objectif matériel au concept de personne – que nous ne sommes pas des organismes, pas des êtres de nature biologique. Quant à l'animalisme que défend Olson, son problème principal, comme on va le voir, c'est que nos capacités mentales

seraient alors non-pertinentes pour déterminer ce que nous sommes : nous serions le même individu sans nos hémisphères cérébraux voire même sans notre cerveau, c'est-à-dire sans la source de notre vie mentale. Les facultés de penser et d'être conscient seraient alors non-essentiels pour les animaux humains, ce qui ne saurait satisfaire. Nous proposerons donc d'explorer l'idée que nous sommes des êtres de nature biologique pour lesquels certaines capacités mentales sont essentielles. La piste que nous allons mettre de l'avant est l'idée que les facultés mentales seraient de nature biologique, du moins chez les animaux.

CHAPITRE I – LA PERSONNE COMME *BUNDLE* D'EXPÉRIENCES

Parfit a défendu de son vivant trois positions différentes sur notre nature ontologique, toutes tributaires de l'approche psychologique lockéenne. Nous allons en discuter chronologiquement afin de montrer l'évolution de sa pensée en réponse à l'ascension de l'animalisme. La position initiale de Parfit, exposée dans le détail dans *Reasons and Persons*, est l'objet de ce chapitre. Elle présuppose, comme la grande majorité des théories de l'identité, que nous, êtres humains, sommes essentiellement des personnes. Pour le premier Parfit, le concept de personne n'est qu'un raccourci langagier pour une collection d'expériences, c'est-à-dire d'événements mentaux,³ causalement interreliés. Notre nature ontologique serait alors celle d'une entité ayant une existence conceptuelle plutôt que réelle.

Dans ce chapitre, nous allons d'abord expliquer la position de Parfit, qu'il nomme le *réductionnisme constitutif*, en utilisant ses textes subséquents pour la clarifier (1.1). Nous allons ensuite examiner certains arguments qu'il donne en sa faveur, ainsi que des objections à celle-ci. En premier lieu, nous allons voir ce qui mène Parfit à croire que le dualisme serait la seule alternative plausible – mais fausse – au réductionnisme, et nous allons défendre que l'animalisme serait également à considérer (1.2). En second lieu, nous allons discuter de la possibilité de la division des hémisphères qui, selon Parfit, fait en sorte que seule une explication réductionniste pourrait rendre compte de l'unité de notre conscience. Nous allons lui opposer des recherches scientifiques récentes qui remettent les prémisses de son argument en doute (1.3). Finalement, nous allons discuter d'une objection fréquemment soulevée contre le réductionnisme de Parfit, à savoir qu'elle aurait pour conséquence de réifier nos expériences (1.4).

3 Parfit utilise les concepts d'événements mentaux et d'expériences interchangeablement, en un sens large qui englobe l'idée d'états mentaux tout comme la persistance dans le temps de croyances et d'autres traits psychologiques (R&P : 211).

1.1 – Parfit I : le réductionnisme constitutif

L'objectif de cette section est d'explicitier le réductionnisme constitutif de Parfit le plus fidèlement possible. Nous allons d'abord examiner comment il justifie son réductionnisme *ontologique*, c'est-à-dire l'idée que nous n'avons pas de réelle existence (1.1.1). Pour lui, un tel réductionnisme implique alors un réductionnisme à propos de notre identité ainsi que de l'importance que nous lui associons (1.1.2). Nous allons ensuite discuter de la portée de ce réductionnisme (1.1.3), ainsi que de l'alternative animaliste que nous allons lui opposer (1.1.4).

1.1.1 – Réductionnisme de l'existence

Parfit identifie deux théories générales dans lesquelles toute théorie sur notre nature ontologique s'inscrit : la *bundle theory*,⁴ d'origine humienne et à laquelle Parfit adhère, et la *ego theory* (Parfit 1987 : 20). Notre tendance naturelle est de nous concevoir conformément à la *ego theory*, c'est-à-dire comme une entité réelle, dont l'existence est distincte de celle d'autres entités réelles. Tout comme le clavier sur lequel je tape, je me conçois comme un objet situé dans le monde et ayant une existence propre. Cette existence consiste en celle d'un sujet qui a des propriétés et qui vit des expériences (sensations, pensées, émotions, etc.), mais qui est distinct de celles-ci. Je ne suis ni mes expériences ni mes attributs, je suis *ce* qui a ces expériences et attributs. Ce qui assure l'unité de ma conscience et de mes expériences durant ma vie, c'est alors le fait que toute expérience dans cette vie appartient au même sujet, *moi*.

Par suite, tant et aussi longtemps que ce sujet réel continue d'exister, je suis le même individu. On dit alors qu'il y a identité *numérique*, par opposition à l'identité *qualitative* qui peut tenir entre deux objets exactement similaires mais numériquement distincts. De plus, puisque la *ego theory* rattache ma survie à celle d'une entité réelle, mon identité est alors déterminée : soit j'existe, soit cette entité est morte ; c'est une question de fait, à propos du réel, et ayant deux réponses possibles. Il n'y a pas de place ni pour des zones grises ni pour des degrés – il n'y a pas de type d'existence entre la vie et

4 Certains traduisent *bundle theory* par « théorie du faisceau ». Cependant, comme cette traduction ne semble pas aller de soi, et que le terme *bundle* est utilisé de manière technique par les auteurs et dans divers contextes, nous allons conserver l'anglais dans ce mémoire.

la mort, ni de plausibilité accordée à l'idée d'une survie partielle ou d'un individu qui serait partiellement moi. Je suis conçu comme un tout unique et indivisible. Quand je me soucie de moi-même, ce qui m'importe le plus c'est la préservation de mon identité numérique, c'est-à-dire de l'existence du sujet réel que je suis, sans quoi je suis réduit à néant. La version la plus connue de la *ego theory* est sans nul doute sa formulation cartésienne, dualiste, selon laquelle le sujet d'expérience est une âme immatérielle ayant une existence distincte du corps – croyance à laquelle Locke souscrivait en privé.

Pour les tenants de la *bundle theory*, à l'inverse, notre existence n'est pas indépendante des éléments qui entrent dans sa constitution. Dans la formulation qu'en donne Parfit, notre existence ne consiste qu'en l'existence d'un corps, d'un cerveau, et d'une série d'événements physiques et mentaux interreliés (R&P : 211) – tout comme l'existence d'une nation ne consiste qu'en l'existence d'un gouvernement, d'un territoire et des individus y vivant ; suivant l'analogie bien connue de Hume (Hume 2007 : 1.4.6, 170 ; SB 261). En effet, nous n'avons jamais directement conscience d'un sujet d'expérience existant indépendamment de nos expériences, comme le remarque Hume : « I never can catch *myself* at any time without a perception, and never can observe any thing but the perception. » (Hume 2007 : 1.4.6, 165 ; SB 252) Il ne semble pas y avoir de substance, de support indépendant qui corresponde à notre idée de l'ego, sinon les expériences elles-mêmes. L'analyse de Locke en venait d'ailleurs à une conclusion similaire malgré les convictions de l'auteur. Notre existence serait donc parasitaire sur celle d'autres entités, plutôt que d'être propre au sujet que nous sommes.

C'est en ce sens que la *bundle theory* est un réductionnisme, plus précisément un réductionnisme *ontologique* comme le désigne Parfit (1999 : 223) : notre existence se réduit à l'existence d'autres entités réelles, faisant de nous des entités n'ayant pas de réalité à proprement parler. Parfit résume son réductionnisme ainsi : « We deny that we are not just conceptually distinct from our bodies, actions, and experiences, but also separately real. » (R&P : 341) Par opposition, la *ego theory* est un non-réductionnisme ontologique, parce qu'elle considère que nous sommes des objets situés dans le monde, qui possèdent chacun une existence propre, indépendante de l'existence d'autres entités. Quel serait alors notre statut ontologique selon la *bundle theory* ? Si nous ne sommes pas de réelles entités, existons-nous, et si oui, en quel sens ?

En niant la réalité de notre existence, la *bundle theory* de Parfit s'apparente au premier abord à un réductionnisme *éliminativiste*. En effet, si notre existence se réduit à celle d'un corps, d'un cerveau, et d'événements qui leur sont reliés, peut-être que la position appropriée à propos de notre statut ontologique consiste à dire que nous n'existons tout simplement pas ; seuls les éléments nous constituant existent. Cependant, Parfit s'oppose à une conception éliminativiste de nous-mêmes, parce qu'elle s'éloigne trop de notre concept usuel de personne. Nations et clubs, quoique ontologiquement réductibles, existent tout de même ; il serait absurde d'affirmer qu'il n'y a pas de telles choses selon lui, et il en va de même pour les personnes (Parfit 1987 : 20). Il vaut donc mieux distinguer entre genres d'entités croit Parfit, selon leur mode d'existence ontologique indépendant ou non par rapport à d'autres entités, plutôt que d'affirmer l'inexistence des entités n'ayant pas d'existence indépendante (Parfit 2007 : 5).

Nous existons certes, mais sous quelle forme existons-nous ? Une réponse possible serait de dire que ce que nous sommes, ce n'est *que* les éléments dont nous sommes constitués. Une personne humaine ne serait alors qu'un corps, un cerveau, ainsi que les expériences qui leur sont associées. C'est ce que Parfit nomme le réductionnisme *identifiant* (Parfit 2007 : 4), qu'il trouve à la fois trop simple, et qui s'éloigne encore une fois de notre concept usuel de personne (R&P : 212). Pour le montrer, il s'appuie sur les analogies suivantes. Malgré qu'une statue de bronze se réduise ontologiquement au bloc de bronze qui la constitue, autrement dit que son existence en soit dérivée et dépendante, il demeure toutefois possible de détruire cette statue sans porter atteinte au bloc, en la faisant fondre par exemple. Cela justifie qu'on n'identifie pas la statue au bloc qui constitue son existence, mais qu'on la considère comme une entité *distincte* (Parfit 2007 : 4). De plus, lorsqu'on réfère à une nation particulière, on ne réfère qu'à elle et non aux éléments qui la constituent ; sans quoi, si son gouvernement ou ses frontières venaient à changer, si une génération de ses citoyens mourrait et était remplacée, on devrait alors dire que la nation en question n'existe plus, ce qui serait absurde (R&P : 472). Il en irait de même pour les personnes.

La position réductionniste appropriée serait alors, aux yeux de Parfit, celle d'un réductionnisme *constitutif*, selon laquelle même si notre *existence* ne consiste qu'en l'existence d'un ensemble d'éléments et s'y réduit, notre *nature*, quant à elle, en demeure

tout de même distincte :

« I am not a series of experiences, but the person who *has* these experiences. A Reductionist can admit that, in this sense, a person is what *has* experiences, or the *subject of experiences*. This is true because of the way in which we talk. What a Reductionist denies is that the subject of experiences is a *separately existing entity*, distinct from a brain and body, and a series of physical and mental events. » (R&P : 223)

Selon un tel réductionnisme, nous ne sommes pas nos expériences, ni notre corps ni notre cerveau, mais nous pouvons affirmer, comme le prétendent les tenants de la *ego theory*, que nous sommes *ce* qui a des expériences, un corps et un cerveau. En d'autres termes, nous ne serions pas le *contenu* du *bundle* qui constitue notre existence, ces éléments ayant une réelle existence, mais bien le *bundle* lui-même, le *contenant*, l'entité abstraite résultant de la collection de ces éléments.

Cette distinction entre l'entité que nous sommes et les éléments dont son existence dépend n'est que conceptuelle, due à notre manière usuelle d'employer le concept de personne : « A Bundle Theorist admits this fact, but claims it to be only a fact about our grammar, or our language. There are persons or subjects in this language-dependent way. » (Parfit 1987 : 20) Nous ne pourrions exister séparément de nos éléments constitutifs, tout comme une nation ne pourrait exister indépendamment de son territoire ou de ses citoyens. Ce n'est que parce que nous faisons couramment la distinction entre nous-mêmes et les éléments qui nous constituent qu'il convient de nous considérer comme une entité distincte d'eux. Ainsi, selon la formulation parfitienne de la *bundle theory*, les personnes humaines ne sont pas des objets réels dans le monde, mais des entités purement conceptuelles, tirant leur nature de notre manière de nous exprimer, et ce même si leur existence dépend d'entités qui sont, elles, réelles.

Cette analyse nous donne une bonne idée de ce que nous sommes fondamentalement pour Parfit. Cependant, c'est là une formulation très générale du réductionnisme ontologique constitutif. En effet, la collection d'éléments auxquels notre existence se réduit supposément – le corps, le cerveau, et les événements physiques et mentaux – *tout* réductionniste à propos de notre existence l'accepterait (R&P : 211). Non seulement le contenu du *bundle* est-il valable pour ceux qui soutiennent un réductionnisme éliminativiste ou identifiant, mais il l'est tout autant pour ceux qui défendent l'idée que nos conditions de persistance sont physiques, psychologiques ou une

combinaison des deux. C'est le cas parce que le *bundle* ainsi spécifié renferme les éléments qui sont effectivement impliqués dans notre existence *actuelle*. Par contre, ce ne sont pas tous les éléments qui s'avèrent nécessaires à notre existence continuée, remarque Parfit : « For example, having the same heart and the same character are not necessary to our continued existence, but they are usually part of what this existence involves. » (R&P : 202) Nous pourrions continuer d'exister, en tant qu'entité *numériquement* identique, même sans certains des éléments qui constituent notre existence présente.

Par conséquent, pour préciser ce que nous sommes aux yeux de Parfit, pour cerner les éléments du *bundle* essentiels à notre existence, il faut nous aventurer sur le terrain de nos conditions de persistance. Parfit considère deux approches à ces conditions : les critères physique et psychologique de l'identité personnelle. Selon le critère d'identité physique, ce qui dans le *bundle* est nécessaire à notre identité numérique et assure la continuation de notre existence, c'est le corps et/ou le cerveau, dépendamment des auteurs. Du côté du critère psychologique tel que le formule Parfit, c'est plutôt notre vie mentale – au sens précis de la série de nos expériences ou encore de nos événements mentaux – qui est nécessaire à notre existence. Pour montrer que l'essentiel, ce serait la continuité de notre vie mentale, Parfit procède en évaluant l'importance relative de chacun des éléments vis-a-vis nos intuitions à propos de notre existence.

L'existence du corps est ce qui est importe le moins : nous savons que nous pouvons survivre à la perte d'un membre voire plusieurs. Même les organes du corps dont la fonction est essentielle à notre survie tel le cœur sont remplaçables, il suffit d'avoir une prothèse artificielle qui sache remplir cette fonction adéquatement. Le problème pour les remplacer n'est donc que technique. Il est même concevable qu'on puisse disposer dans un futur proche des moyens de survivre si seul notre cerveau continuait d'exister : « Receiving a new skull and a new body is just the limiting case of receiving a new heart, new lungs, new arms, and so on. » (R&P : 253) Quoique notre corps affecte le contenu de nos expériences, la perte d'un corps particulier ne peut changer le fait que les expériences qu'on vit sont nôtres (Parfit 2008 : 180), et que nous continuons donc d'être la même entité. Un critère d'identité physique plus plausible n'exigerait donc que l'existence continuée du cerveau – ou plus précisément de *suffisamment* du cerveau pour soutenir la

vie et la conscience, puisqu'on sait que des individus survivent à des lésions cérébrales.

Nos intuitions nous inclinent à croire qu'on ne pourrait exister sans notre cerveau, surtout si l'on est physicaliste. Nagel, quant à lui, affirme même que nous sommes essentiellement notre cerveau, plutôt qu'une entité conceptuelle qui en dépendrait pour exister (Nagel 1986 : 40). Toutefois, la nécessité du cerveau ne demeure qu'hypothétique, puisque notre concept de *soi* ou de *personne* est indépendant de tout substrat, et que nos connaissances semblent insuffisantes pour l'établir. Selon Parfit, si l'on croit que l'existence de notre cerveau est essentielle à la nôtre, ce n'est que parce qu'elle assure la continuité de notre vie mentale :

« Why should the brain be singled out in this way? The answer must be: 'Because the brain is the carrier of psychological continuity, or Relation R'. If this is why the brain is singled out, the continuity of the brain would not matter when it was not the carrier of Relation R. The continuity of the brain would here be no more important than the continuity of any other part of the body. And the continuity of these other parts does not matter at all. » (R&P : 284)⁵

Ce serait donc avant tout la série de nos événements mentaux qui serait selon Parfit essentielle à notre existence. Le cerveau n'aurait d'importance que comme support de notre vie mentale et cause normale de sa continuité. Cela s'aligne avec la définition lockéenne de la personne que Parfit reprend, selon laquelle pour être une personne, il faut que l'entité soit consciente de son identité dans le temps (R&P : 202) – ce qui requiert l'identité psychologique. La persistance de notre cerveau sans celle de notre vie mentale ne saurait alors assurer notre existence ; et même si elle s'avérait nécessaire, elle ne saurait selon Parfit être suffisante.

La question qui se pose alors, c'est de savoir si la série de nos expériences pourrait subsister sans notre cerveau. Cet organe est-il, comme les autres, hypothétiquement remplaçable par une prothèse pouvant remplir ses fonctions ? Parfit admet qu'il est plausible que notre vie mentale requiert un cerveau humain comme support matériel, voire une copie exacte du nôtre si elle n'est physiquement réalisable que d'une seule manière ; mais doit-il être le *même* ? Parfit est ambivalent à ce sujet, étant donné nos connaissances insuffisantes pour trancher.

Il croit cependant que si notre vie mentale pouvait continuer d'exister sans notre

⁵ Dans *R&P*, Parfit nomme « Relation R » la relation de continuité psychologique requise pour l'identité personnelle. Elle n'inclut cependant pas la condition d'*unicité* que nécessite l'identité.

cerveau, alors elle seule serait nécessaire et suffisante à notre identité. C'est pourquoi, pour lui, ce que nous sommes essentiellement et qui forme le contenu du *bundle* dans sa formulation plus spécifique, ce n'est que la série de nos expériences :

« [T]here are long series of different mental states and events – thoughts, sensations, and the like – each series being what we call one life. Each series is unified by various kinds of causal relation, such as the relations that hold between experiences and later memories of them. Each series is thus like a bundle tied up with string. » (Parfit 1987 : 20)

Voilà qui nous fixe donc sur la position de Parfit quant à notre nature ontologique. Nous sommes essentiellement des entités de nature conceptuelle dont l'existence dépend de celle de notre vie mentale.

1.1.2 – Réductionnismes de l'identité et de son importance

Attardons-nous maintenant sur ce qui pour Parfit assure notre identité dans le temps. Dans le passage que nous venons de citer, non seulement Parfit décrit-il ce qui constitue l'existence du *bundle* conceptuel que nous sommes, il explique également ce qui en constitue l'*identité*. Alors que notre existence dépend des éléments constituant notre vie mentale, à savoir nos événements mentaux, notre identité, quant à elle, dépend des relations entre ces éléments. En d'autres termes, ce qui fait des éléments nous constituant les éléments d'*un seul et même* individu, ce sont les relations causales entre ceux-ci – entre souvenirs, intentions, désirs, et traits psychologiques passés et présents, si on adopte le critère d'identité psychologique.⁶ Par suite, cet individu persiste dans le temps tant que de telles relations raccordent ces expériences ensemble en une série, et que de nouvelles expériences se rajoutent à cette série.

Quelles sont alors ces relations qui constituent notre identité et sa persistance ? Comme le souligne Parfit, il y a des relations causales qui assurent la persistance de nos événements mentaux dans le temps, tel le fait que j'aie encore le souvenir d'un événement ayant eu lieu il y a dix ans, ou encore que mes intentions soient maintenues jusqu'à temps que je les réalise ou les abandonne. Parfit parle alors de connexions psychologiques directes et les regroupe sous le terme de *connectivité* (*connectedness*) psychologique.

⁶ J'ometts ici la discussion de l'*unicité* comme condition additionnelle pour l'identité. Nous en ferons mention plus loin, quoique cette question ne soit pas centrale dans ce mémoire, puisqu'il est avant tout à propos de notre nature ontologique plutôt que de l'identité.

D'un jour à l'autre, un bon nombre de telles connexions directes sont conservées, suffisamment pour qu'une personne demeure la même⁷, ce que Parfit nomme une connectivité *forte*. Toutefois, malgré que tout au long de notre existence nous maintenions normalement une forte connectivité par rapport au jour précédent, à chaque jour certaines connexions (souvenirs, intentions, etc.) précédemment maintenues se perdent. Si l'on remonte suffisamment dans le passé, nous ne serons pas en forte connectivité avec notre moi passé, nos croyances, intentions etc. ayant significativement changé, malgré que nous soyons le même individu. La connectivité ne peut donc expliquer à elle seule le maintien de notre identité dans le temps.

Il nous faut donc une relation transitive, capable de valoir pour l'entièreté de notre existence malgré nos changements psychologiques. Or, les multiples chaînes de connexions psychologiques directes qui se chevauchent durant la vie d'une personne forment ensemble la relation générale que Parfit nomme la *continuité* psychologique (R&P : 206), et c'est elle qui tisse nos expériences en une série persistante. Si maintenant nous ne sommes pas directement psychologiquement connectés à nous-mêmes lorsque nous avons trois ans, parce que nos croyances, intentions etc. ont changé, nous y sommes tout de même connectés indirectement par cette relation de continuité qui atteste de la causalité à l'œuvre dans l'entièreté de notre vie mentale. Des expériences désormais oubliées ont causé chez nous des dispositions à l'origine de nos intentions d'agir actuelles, par exemple. La persistance de notre identité s'explique donc par le maintien de la continuité psychologique à travers notre vie mentale, c'est-à-dire par les relations causales ponctuelles entre nos événements mentaux, sous-jacentes à la relation générale de continuité.

Ainsi, notre identité *ne consiste qu'en* l'existence de relations causales entre les éléments du *bundle* conceptuel que nous sommes, et notre persistance dans la continuité de l'existence de telles relations (Parfit 2007 : 6). C'est donc là de nouveau un réductionnisme constitutif⁸. Cette fois-ci par contre, ce réductionnisme a un sens différent que précédemment : alors que plus tôt c'était notre *existence* qui était réduite à l'existence

7 Parfit affirme qu'on ne peut déterminer précisément combien de connexions directes – c'est-à-dire quel degré de connectivité – sont suffisantes pour l'identité (R&P : 206).

8 Selon Parfit, tout critère d'identité, qu'il soit physique ou psychologique, est une manière réductionniste de rendre compte de nos conditions de persistance (R&P : 210).

des éléments du *bundle*, c'est maintenant le fait de notre *identité* (et de sa persistance) qui est réduit aux faits de continuité psychologique qui relient ces éléments en un *bundle*.

Par conséquent, tout comme selon le réductionnisme constitutif notre existence n'était pas indépendante de l'existence d'autres entités réelles, notre identité n'est pas non plus un fait indépendant d'autres faits :

« According to Constitutive Reductionism, the fact of personal identity is distinct from these facts about physical and psychological continuity. But, since it just consists in them, it is not an independent or separately obtaining fact. It is not a further difference in what happens. » (Parfit 2007 : 6)

Pour reprendre la formule utilisée plus tôt, le fait que nous sommes identiques à nous-mêmes serait parasite sur les faits à propos des relations causales entre nos expériences. Et comme c'était le cas précédemment pour notre existence, cette entière dépendance de notre identité sur d'autres faits n'implique pas que c'est alors un fait à éliminer ou à identifier aux faits qui le constituent. S'il existe des personnes, il serait absurde de dire qu'il n'y a toutefois pas de faits à propos d'eux. En poursuivant avec l'analogie humienne, il y a tout de même des faits à propos de nations même si ceux-ci se réduisent à un ensemble de faits à propos d'un territoire et d'individus (Parfit 2007 : 7). Un fait non-indépendant demeure donc tout de même distinct des faits qui le constituent.

De plus, comme c'était le cas pour notre nature ontologique, notre identité est alors un fait de nature différente des faits qui le constituent. La réduction constitutive, qui plus tôt faisait de nous des entités conceptuelles plutôt que réelles, fait de notre identité un fait conceptuel plutôt que réel :

« [I]f one fact just consists in certain others, the first fact is not an independent or separately obtaining fact. [...] [I]n relation to those other facts, the first fact is merely conceptual. » (Parfit 2007 : 32)

L'identité est un fait à propos du *bundle* conceptuel que nous sommes, plutôt que directement à propos des éléments qui constituent notre existence et des relations entre eux. Un tel fait, même s'il n'est pas indépendant des faits à propos de nos constituants, a tout de même une raison d'être, une pertinence conceptuelle. Qu'une nation déclare la guerre, c'est un fait entièrement dérivé de faits à propos de gestes posés par seulement quelques individus appartenant à cette nation, mais qui engagent l'ensemble, ce que le fait à propos de la nation exprime clairement et simplement. On pourrait par contre en exprimer pleinement le contenu en énonçant l'ensemble des faits dont il est constitué. Ce

n'est donc pas un *further fact*⁹ par rapport aux faits qui le constituent ; ainsi en est-il de l'identité des personnes (R&P : 210). Affirmer ou nier l'identité d'une personne n'implique rien de plus que les faits auxquels cette identité se réduit, en l'occurrence les faits de continuité psychologique. Le fait de l'identité n'offre donc pas de contenu nouveau hormis à propos de notre façon de parler, autrement dit à propos de notre langage et du schème conceptuel qu'il met en œuvre (Parfit 2007 : 7). Sa seule utilité est de faciliter l'expression d'un ensemble de faits réels.

Ensuite, le réductionnisme constitutif à propos des faits entraîne un réductionnisme additionnel, cette fois-ci à propos de l'*importance* de ces faits. Selon Parfit, puisque les faits conceptuels n'expriment rien de plus que des ensembles de faits réels, ce sont les seuls faits réels qui importent : « If one fact just consists in certain others, it can only be these other facts which have rational or moral importance. » (Parfit 2007 : 28) C'est ce qu'il nomme le *réalisme de l'importance* :

« What matters are the facts about the world, given which some concept applies. If the facts about the world have no rational or moral significance, and the fact that some concept applies is not a further difference in what happens, this conceptual fact cannot, I believe, be significant. » (Parfit 2007 : 33)

L'importance accordée à des faits conceptuels n'est que dérivée de l'importance des faits réels dont ils sont constitués, et y est donc réductible. Ainsi, puisque le fait de notre identité se réduit à des faits de continuité psychologique, ce sont alors les seconds qui importent et non notre identité. Ce qui importe donc, c'est seulement la continuité de la série de nos expériences. Les faits conceptuels n'ont par eux-mêmes aucune importance, c'est-à-dire qu'ils n'ont qu'une importance indirecte étant donné par exemple le rôle légal qu'on leur accorde, notamment à l'identité pour trancher des questions de propriété et d'héritage (Parfit 2007 : 33).

Dans *Reasons and Persons*, Parfit intégrait ce réductionnisme à ceux exposés précédemment sous la bannière *du* réductionnisme, ce qu'il a explicitement admis comme une erreur, parce que plusieurs réductionnistes à propos de notre existence et notre

9 Parfit utilise l'expression *further fact* en un sens précis dans *Reasons and Persons*. Si l'on défend l'idée que notre identité ne se réduit pas à des faits de continuité physique et/ou psychologique, alors on croit selon lui que notre identité est un *further fact*. C'est ce qu'auraient en commun toutes les positions non-réductionnistes et ce qui les distingue du réductionnisme. Montrer qu'il n'y a pas de *further fact* reviendrait alors à montrer que le réductionnisme est vrai. L'expression symbolise donc le débat, c'est pourquoi nous conservons l'anglais ici.

identité ne partagent pas celui touchant son importance (Parfit 2007 : 27). Le réductionnisme à propos de l'importance serait indépendant des autres.

En somme, le réductionnisme de Parfit sous-jacent à sa *bundle theory* est en fait un triple réductionnisme. D'abord, c'est un réductionnisme à propos de notre existence. Celle-ci qui ne consiste essentiellement qu'en l'existence de nos expériences, quoiqu'en pratique elle fait également intervenir un corps, un cerveau et des événements physiques malgré qu'ils ne soient pas essentiels. Cette réduction de notre existence à celle d'un ensemble d'éléments fait de nous des entités de nature conceptuelle plutôt que réelle. Ensuite, c'est un réductionnisme à propos des faits de notre identité et de notre persistance. Ces faits ne consistent que dans la présence de faits à propos des relations causales entre nos expériences, notamment la relation de continuité psychologique. Finalement, c'est un réductionnisme à propos de l'importance de notre identité et de sa persistance. Ces dernières ne consistant que dans des faits à propos de relations, toute l'importance appartient donc aux relations causales elles-mêmes, aux faits réels. Le réductionnisme de Parfit fait donc de nous des entités irréelles sans importance.

1.1.3 – Portée du triple réductionnisme

La combinaison de ces réductionnismes a d'importantes conséquences. D'abord, si nous, personnes, ne sommes que des entités conceptuelles, et que le fait de notre identité dans le temps n'est que conceptuel, alors toute description impliquant des personnes ne nous dit rien de plus que les descriptions qu'on peut donner à propos d'expériences et des relations entre elles, hormis à propos de notre schème conceptuel. Cela conduit Parfit à émettre l'étonnante déclaration suivante : « Though persons exist, we could give a *complete* description of reality *without* claiming that persons exist. » (R&P : 212) Comprenons-là ainsi : pour fournir une description complète du réel, il nous faut seulement rendre compte d'entités et de faits réels, parce que seuls ceux-ci font une différence. Comme les entités et les faits conceptuels n'expriment aucun contenu de plus que les entités et les faits réels dont ils sont constitués et n'ont donc pas d'importance rationnelle ou morale additionnelle, nous n'avons pas besoin de les mentionner : « [since they] are two ways of describing the same fact[s], a complete description need not make *both* claims. » (R&P : 212) Les entités et faits conceptuels seraient déjà sous-entendus

dans une description impersonnelle de la réalité. *Nous* sommes superflus.

Une autre conséquence du réductionnisme de Parfit est que la majorité d'entre nous arborons de fausses croyances à propos de notre nature ontologique (R&P : 217). La science actuelle nous donne de bonnes raisons de croire que la *ego theory*, plus précisément sa formulation cartésienne selon laquelle nous serions des âmes immatérielles, est fautive : « Given the advances in psychology and neurophysiology, the Bundle Theory may now seem to be obviously true. It may seem uninteresting to deny that there are separately existing Egos » (Parfit 1987 : 21). Et pourtant, la plupart de ceux qui croient en la fausseté de la *ego theory* maintiennent tout de même des croyances qui en sont dérivées, remarque Parfit (1987 : 21). Son objectif premier est donc, plutôt que de réfuter la version cartésienne de la *ego theory*, de nous montrer les conséquences de l'adoption du réductionnisme pour que nous fassions nôtres des croyances cohérentes alignées sur la *bundle theory*. Malgré notre résistance, « We can believe the truth about ourselves. » (R&P : 280)

Tel que mentionné plus haut, une croyance propre à la *ego theory* consiste en ce que notre identité est déterminée et une question « tout-ou-rien » : soit nous demeurons numériquement identique à nous-mêmes, soit nous cessons d'exister ; la question de savoir si on survit a toujours une réponse et n'offre que deux possibilités. Cependant, si notre existence se réduit à l'existence d'un ensemble d'expériences reliées, il serait concevable que seulement un sous-ensemble de ces expériences soit maintenu. Notre survie pourrait alors n'être qu'une question de degré, et même s'avérer indéterminée dans certains cas. Mais comme notre identité n'est qu'un fait conceptuel selon Parfit, n'exprimant pas une différence dans le réel en dehors des relations de continuité sous-jacentes, cette indétermination ne serait pas problématique. Nous expliquerons cela plus en détail dans la prochaine section (1.2).

Finalement, si le réel est pleinement descriptible et explicable sans mentionner les personnes, alors toute forme d'explication doit pouvoir être reformulée sans dépendre du concept de personne ou de l'identité de telles entités. L'unité de la conscience et de la vie d'un individu, notamment, doit pouvoir s'expliquer sans dire que nos expériences sont unifiées par leur appartenance à une personne ou un sujet d'expérience. Les relations causales entre celles-ci offrent une bonne partie de la réponse ; nous y reviendrons (1.3).

De plus, notre agir et notre manière de parvenir à des décisions rationnelles et morales doivent également pouvoir être libérées de la notion de personne. En effet, nous avons la croyance que notre souci de soi et que la responsabilité de nos actes se fondent sur notre identité : c'est parce que les événements m'arrivent à *moi*, et que mes actions sont les miennes, que je m'en soucie et que j'en suis responsable. C'était là une idée centrale chez Locke. Mais si Parfit n'a pas tort, et que notre identité n'est qu'un fait conceptuel, alors le souci pour soi et les autres ainsi que l'agir rationnel et moral doivent être expliqués autrement, sans dépendance à notre identité, mais plutôt en rapport à des faits réels. En effet, sans le fait réel de l'identité sur lequel s'appuyer, « the principle of *self-interest* has no force. » (Parfit 1971a : 26 – c'est moi qui souligne) Nous n'entrerons pas dans le détail de cette question dans ce mémoire.

Bref, le triple réductionnisme de Parfit a de vastes implications, notamment sur notre manière de décrire le réel, sur notre conception de nous-mêmes, et sur le fondement de la responsabilité ainsi que de l'agir rationnel et moral.

1.1.4 – L'animalisme comme alternative au réductionnisme

Le point de vue qui sera discuté dans ce mémoire en réponse au réductionnisme qui vient d'être présenté, c'est que Parfit se méprend par rapport à notre nature ontologique fondamentale. La *bundle theory* ne serait pas, contrairement à ce qu'il prêche, la « *true view* » (R&P : 274) à propos de nous-mêmes. *Nous* pourrions être des entités réelles, situées dans le monde, sans qu'on doive toutefois postuler l'existence d'entités additionnelles telles les âmes immatérielles.

Parfit admet certes que sa position puisse comporter des erreurs ; cependant, il avance ceci : « my main claim is that persons are like nations, not Cartesian Egos. If this claim is true, it would not be undermined by my mistakes. » (R&P : 275) Ma réponse est la suivante. Même si nous acceptions cette thèse ; même si nous devons être réductionnistes à propos des personnes, c'est-à-dire à propos de leur nature, de leur existence, de leur identité, de leur persistance ainsi que de leur importance ; même si nous pouvions donner une description complète du réel sans les mentionner ; nous n'avons quand même pas à être réductionnistes à propos de nous-mêmes et conclure que nous ne sommes que des entités conceptuelles, parce qu'il y a une alternative plus

plausible : à savoir que nous ne sommes pas fondamentalement des personnes humaines, mais des animaux humains, qui, *eux*, sont des entités réelles, ayant une existence indépendante de leurs constituants, et dont la nature correspond aux exigences de la *ego theory*. L'animalisme pourrait donc être compris comme une version de la *ego theory* alternative au dualisme cartésien, qui nous conçoit comme des objets matériels plutôt qu'immatériels.

Ainsi, on pourrait accorder à Parfit son réductionnisme ontologique à propos des personnes, tout en demeurant non-réductionniste à propos de ce que nous sommes fondamentalement. La *bundle theory* serait compatible avec la *ego theory*, mais leur objet serait différent. Les personnes seraient alors des entités conceptuelles dépendant de l'existence d'entités réelles, les animaux humains dans notre cas – ce qui n'empêche pas la possibilité qu'il puisse exister des personnes qui ne soient pas des animaux humains, d'origine extraterrestre ou artificielle. Si nous n'avons pas tort, l'erreur principale de Parfit serait donc de suivre la tradition lockéenne dans son association du concept de personne à *soi*, à ce que nous sommes essentiellement.

Par suite, si effectivement nous sommes essentiellement des animaux, et que ceux-ci sont des entités réelles, alors certaines des assertions de Parfit quant aux implications de son réductionnisme de la personne seraient à réviser. Avant tout, même s'il était possible de fournir une description impersonnelle du réel, elle ne pourrait être complète sans *nous* y inclure en tant qu'animaux. Puisque l'existence d'entités réelles ne serait pas réductible à celle d'un ensemble d'éléments constitutifs, une description de ces seuls éléments ne nous inclurait pas. De plus, il ne serait pas non plus nécessaire de fournir une explication de l'unité de la vie et de la conscience indépendante du sujet d'expérience si le sujet d'expérience *est* l'animal. D'ailleurs, comme on va le voir plus loin, la raison principale que Parfit a de croire que la conscience et l'animal sont séparables, à savoir la réelle possibilité de division de la conscience chez un même animal, serait fausse.

Surtout, les croyances intuitives que nous avons à propos de nous-mêmes et de notre identité ne seraient pas à rejeter. Notre identité numérique en tant qu'animal humain ne serait pas qu'un fait conceptuel sans importance : si je ne persiste pas, c'est que l'animal que je suis est mort. Il n'y aurait ni survie partielle, ni survie ontologiquement

indéterminée, la mort de l'organisme étant absolue, définitive et sans recours. Par conséquent, nos croyances que nous sommes des entités réelles, que notre identité importe, qu'elle est déterminée, et qu'elle ne peut tenir à un degré partiel, pourraient être pleinement explicables et justifiées par une position animaliste. De plus, le souci de soi et des autres pourrait alors être compris comme un souci à propos d'une entité réelle, justifiant le fondement classique de l'agir rationnel et moral sur la *self-interest theory* que Parfit rejette.

Ainsi, si l'animalisme était effectivement une version de la *ego theory*, un des grands avantages qu'il aurait sur le réductionnisme de Parfit serait de pouvoir fournir une conception de notre nature ontologique qui est compatible avec nos intuitions et qui saurait en rendre compte, plutôt que de les balayer comme de vulgaires artefacts de notre vanité et de notre peur irrationnelle de la mort.

1.2 – La *ego theory* peut-elle être défendue ?

Dans la section précédente, nous avons présenté en détail le réductionnisme ontologique de Parfit selon lequel nous ne serions pas des entités réelles puisque notre existence se réduirait à celle des éléments qui entrent dans notre constitution. Nous ne serions alors que le *bundle* conceptuel rassemblant ces constituants. Dans la présente section, nous allons examiner la théorie rivale à propos de notre nature ontologique, la *ego theory*.

Parfit croit que toute position non-réductionniste à propos de nous-mêmes serait à rejeter. Nous ne sommes pas entrés jusqu'ici dans le détail de ses raisons, hormis pour mentionner que nous ne sommes pas immédiatement conscients que nous sommes une entité réelle distincte de notre corps, de notre cerveau et de nos expériences. L'argument principal de Parfit contre la *ego theory* consiste à montrer que la seule version viable de celle-ci serait le dualisme cartésien, mais que comme les données empiriques indiquent qu'elle serait fautive, la seule option restante et valable serait le réductionnisme. C'est l'objet de la sous-section qui suit (1.2.1).

Dans les sous-sections subséquentes, nous allons examiner s'il y aurait une alternative au dualisme cartésien que Parfit ne considère pas, à savoir l'animalisme, qui

pourrait être une version plausible de la *ego theory*. Pour ce faire, il nous faut avant tout voir s'il est cohérent de considérer que les animaux humains échappent au réductionnisme constitutif de Parfit (1.2.2), sans quoi l'animalisme ne pourrait prétendre être une version de la *ego theory*. Nous évaluerons ensuite si les croyances à propos de notre identité qui découlent de l'animalisme correspondent à celles de la *ego theory* (1.2.3). Les arguments directement en faveur de l'animalisme feront plutôt l'objet des chapitres suivants (2.1 et 3).

Parfit développe également un deuxième argument majeur selon lequel la *ego theory* ne serait pas recevable. L'apparente possibilité de la division de la conscience d'un individu en deux consciences distinctes, supportée par des recherches scientifiques, indique qu'une entité réelle ne pourrait être ce qui unifie la conscience, et qu'alors la seule explication plausible de son unité serait réductionniste. Cet argument sera l'objet d'une section ultérieure (1.3).

1.2.1 – Le dualisme cartésien, seule version plausible de la *ego theory* ?

Un des arguments principaux de Parfit en faveur du réductionnisme et de la *bundle theory* est l'expérience de pensée des *spectres de connectivité psychologique et/ou physique* que nous allons présenter. Elle montre selon lui que les assertions à propos de notre existence et de notre identité sont liées ; autrement dit, que les réductionnismes de l'existence et de l'identité vont de pair. Ce ne serait que si nous avions une existence séparée ou indépendante des éléments du *bundle* nous constituant – d'un corps, d'un cerveau et d'événements physiques et mentaux – que notre identité serait toujours déterminée, c'est-à-dire qu'il y aurait toujours une réponse à la question de savoir si un individu est nous ou non (R&P : 240). En d'autres termes, ce ne serait que si nous étions une entité de nature réelle plutôt que conceptuelle que notre identité serait une affaire tranchée, parce qu'elle dépendrait alors de l'existence continuée ou non de cette entité : je serais *moi* tant que l'entité réelle que je suis continue d'exister. Pour Parfit, cela signifie que la seule version plausible de la *ego theory* est celle du dualisme cartésien selon laquelle nous sommes une âme, distincte du corps et siège de notre vie mentale. Voyons pourquoi.

Alors que notre identité est déterminée dans des cas réels, l'expérience de pensée

des spectres met en scène des cas problématiques, peut-être impossibles, pour contester nos croyances intuitives à propos de nous-mêmes. Commençons par celle du spectre psychologique (R&P : 231), qui présuppose que nous adoptions un critère d'identité psychologique. L'expérience nous place sur une table d'opération, entre les mains d'un chirurgien s'appêtant à remplacer une à une toutes nos expériences – mémoires, intentions, etc. – par celles d'un inconnu. Le processus graduel de l'opération produit un *spectre* de cas couvrant tous les degrés possibles de connectivité psychologique, partant de cas présentant un très fort degré de connectivité avec nous, n'ayant perdu que quelques uns de nos souvenirs et de nos croyances, et allant à des cas n'ayant qu'un degré de connectivité très faible, leur vie mentale étant presque entièrement celle de l'inconnu. Pourvu que nous ayons adopté le critère psychologique, il nous semble intuitif que nous survivrions à l'opération dans les cas où il y a forte connectivité et où la majorité de nos expériences sont conservées, alors que la personne résultante ne serait pas nous dans les cas de faible connectivité.

Les cas problématiques sont ceux au milieu du spectre. Un individu avec autour de 50 % de nos expériences et le reste provenant d'un inconnu, serait-il nous, ou pas ? Si l'on reformule la question, à quel moment durant l'opération ne serions-nous plus nous-mêmes ? Même si cela semble difficile à déterminer, nous croyons intuitivement, en accord avec la *ego theory*, qu'il y a une réelle différence entre le fait que l'individu soit nous ou non, et qu'il y a donc une réponse claire à la question. Cependant, l'opération étant graduelle, cette exigence d'une identité déterminée implique qu'à un moment de l'opération, la soustraction d'une expérience de plus ferait de la personne résultante quelqu'un d'autre que nous. La perte d'une seule connexion psychologique tel un souvenir – un changement infime, presque trivial – ferait basculer notre identité numérique de la vie à la mort. Cela est difficile à croire, selon Parfit (R&P : 231).

L'alternative, selon laquelle aucun changement infime ne pourrait affecter de la sorte notre identité, n'est par contre pas davantage plausible, parce qu'elle implique qu'à force de tels changements, nous demeurerions identiques à nous-mêmes sans aucune de nos connexions psychologiques. On pourrait certes croire que l'expérience de pensée montre en fait que le critère d'identité psychologique n'est pas approprié et que nous sommes le même individu à tout moment de l'opération, même sans aucune connexion

psychologique, parce que le critère correct de notre identité est plutôt physique. Cependant, l'expérience de pensée peut être reproduite pour le critère physique – les cellules de notre cerveau étant graduellement remplacées cette fois – avec les mêmes implications. Si l'on croit que notre identité est déterminée, on doit alors admettre que la perte d'une cellule de plus à un moment de l'opération nous changerait complètement au point où nous cesserions d'exister en tant que le même individu. On ne peut ainsi échapper au problème en changeant de critère. Nous n'avons donc que deux options : soit nous acceptons le réductionnisme de Parfit, soit nous admettons qu'il y a une nette démarcation dans les spectres de connectivité, aussi invraisemblable soit-elle.

L'expérience de pensée est analogue au paradoxe sorite, qui nous mène fallacieusement à conclure qu'un seul grain de sable est un tas, parce que chaque étape où l'on enlève un grain de plus du tas initial est un changement trop minime pour qu'on affirme légitimement qu'il affecte la nature de l'objet. Ce paradoxe exploite le fait que le concept de tas est vague, avec une ligne de démarcation floue (R&P : 232). Il réfère d'ailleurs à une entité conceptuelle : un tas de sable n'a pas d'existence indépendante des grains qui le constituent. Dans ce cas, on peut régler le problème par une stipulation fixant le nombre de grains requis pour former un tas. Une telle précision arbitraire appliquée au concept de personne ne réglerait cependant pas le problème qui nous préoccupe, parce qu'elle ne saurait faire de nous des entités réelles ayant une identité *intrinsèquement* déterminée de par leur nature. Elle ne saurait valider la *ego theory*. Peter Unger avance même que s'il est paradoxal de croire que le retrait d'une connexion psychologique ou d'une cellule ferait toute la différence entre notre survie ou notre mort, c'est parce qu'il n'y avait en fait jamais quelqu'un là (Unger 1979 : 206).

Si une stipulation ne saurait expliquer la présence d'une démarcation nette dans le spectre de continuité et attester de notre identité déterminée, que faudrait-il alors ? Selon Parfit, la seule justification admissible pour croire qu'il y a chez nous un ensemble essentiel à notre identité de connexions psychologiques et/ou de cellules du cerveau, serait que nous sommes des entités réelles pour lesquelles porter atteinte à cette part essentielle signifierait la mort plutôt qu'une survie partielle :

« We cannot defensibly believe that our identity involves a further fact, unless we also believe that we are separately existing entities, distinct from our brains and bodies. And we cannot defensibly believe that our identity

must be determinate, unless we believe that the existence of these separate entities must be all-or-nothing. » (R&P : 240)

Or, si nous étions en mesure d'effectuer les opérations chirurgicales que mettent en scène les expériences de pensée des spectres de connectivité, une position selon laquelle nous sommes des entités réelles dont la survie est une question tranchée serait empiriquement vérifiable (R&P : 237). En effet, si cette position était vraie, la soustraction d'une expérience ou d'une cellule essentielle occasionnerait alors la mort de l'entité porteuse de l'identité numérique, correspondant à un changement radical dans la vie mentale de l'individu résultant, voire à sa mort physique. C'est pourquoi, si nous croyons que nous sommes une telle entité réelle, il serait justifiable de croire que notre identité ne se réduit pas aux seuls faits de continuité, mais dépend d'un *further fact*, à savoir la vie de cette entité.

Par contre, Parfit juge qu'une position selon laquelle nous aurions une identité déterminée malgré que nous n'avons pas d'existence indépendante de nos constituants n'est pas défendable (R&P : 239). En effet, on ne pourrait *jamais* avoir de données empiriques quant à l'emplacement de la démarcation dans le spectre de connectivité si nous ne sommes pas une entité réelle dont la mort en serait la preuve observable. Dans le cas du spectre physique, la personne croirait toujours être la même puisque la continuité psychologique serait maintenue dans tous les cas ; alors que dans le cas du spectre psychologique, nous n'aurions aucune raison de nous fier aux croyances de la personne dont la vie mentale est modifiée : « In this kind of case, who someone is cannot be shown by who he thinks he is. » (R&P : 238). Ainsi, si nous ne croyons pas que nous sommes une entité ayant une existence indépendante, il semblerait que nous n'ayons pas de raison suffisante de croire que notre identité ne se réduit pas aux seuls faits de continuité psychologique et/ou physique. Et si l'existence de cette entité n'est pas alors une question « tout-ou-rien », puisque les faits de continuité peuvent tenir à divers degrés, rien ne semblerait pouvoir justifier qu'une différence en apparence triviale telle la perte d'une seule connexion psychologique ou d'une cellule fasse de nous quelqu'un d'autre (ou un cadavre).

Puisque Parfit suit Locke dans l'association de l'identité personnelle à la vie mentale, il considère que la seule entité réelle qui puisse remplir de façon plausible les

exigences non-réductionnistes pour une identité déterminée, c'est l'âme immatérielle. Si nous étions des âmes, la perte de connexions et/ou de cellules essentielles causerait la mort de l'âme ou la fin de sa capacité à interagir avec le corps, de sorte que la psychologie de l'animal humain résultant, s'il survit, serait complètement transformée (R&P : 237). La seule alternative plausible au réductionnisme serait donc le dualisme cartésien. L'animal n'est même pas considéré comme candidat pour cette entité réelle dans *Reasons and Persons*. Il n'est d'ailleurs pas si évident que l'animal soit une entité ayant une existence indépendante de son corps et de son cerveau, contrairement à l'âme – ce sera là l'objet de la prochaine sous-section (1.2.2).

Cependant, les données que nous avons à propos de nous et du réel ne supportent pas l'existence des âmes et tendent plutôt à montrer le contraire. Nous avons déjà discuté du fait que nos expériences ne nous rendent pas directement conscients de l'existence d'une telle entité séparée de notre corps ; notre conscience est donc au moins conceptuellement distincte de toute substance qui en serait le support. De plus, Parfit avance que si nous étions des âmes, la résurrection serait alors possible et aurait dû laisser des traces tangibles (R&P : 227). Des personnes auraient pu se remémorer des faits à propos de leurs vies passées qui auraient pu être vérifiés, comme la présence d'objets enfouis ou de tombeaux qu'il serait possible de dater. Mais nous ne disposons d'aucune confirmation de la sorte.

Ensuite, si nous étions des âmes distinctes de notre corps, nous devrions également pouvoir constater qu'un ensemble de changements physiques, de maladies et de blessures au cerveau, n'affecteraient nullement la personne, alors que tout changement additionnel briserait le lien entre le corps et l'âme, changeant complètement la personne, comme le suggère l'expérience des spectres. Cependant, nous avons de nombreuses preuves que notre vie mentale est affectée par des accidents ou maladies du cerveau, indiquant que son support ne serait pas l'âme mais plutôt le cerveau. Finalement, les données scientifiques suggéreraient que le porteur de la continuité psychologique est divisible, ce qui semblerait éliminer la possibilité que l'identité de la personne soit rattachée à une entité réelle indivisible. Nous y reviendrons en détail plus loin (1.3). En somme, nos connaissances semblent éliminer le dualisme : « if we have no reasons to believe that such entities exist, we should reject this belief. » (R&P : 224) La seule option

crédible serait le réductionnisme, qu'on devrait alors adopter. L'échec du dualisme implique donc selon Parfit qu'il faut rejeter la *ego theory*, et avec elle les croyances intuitives que nous avons à propos de notre identité.

Par conséquent, parce que nos croyances intuitives sont inconsistantes dans les cas qui se situent au centre des spectres de connectivité, on devrait plutôt conclure que notre identité est indéterminée dans de tels cas, selon Parfit (R&P : 240). La question de savoir si c'est nous qui survivons ou non n'aurait tout simplement pas de réponse. Cependant, cette indétermination n'est pas problématique : s'il était en théorie possible de connaître tous les faits à propos des relations sous-jacentes à l'identité, la question de l'identité numérique serait alors vide et sans importance (R&P : 233), parce qu'elle ne serait pas un fait additionnel à propos du réel. Comme on l'a vu, selon le réductionnisme de Parfit, le fait de notre identité n'est que conceptuel ; or, seuls les faits réels, à propos des entités réelles et des relations entre elles, font une différence et ont une importance. Affirmer que la personne résultante est nous ou quelqu'un d'autre, ce ne seraient là que des descriptions équivalentes : « These are not here two different possibilities, one of which must be true. These are merely two different descriptions of the very same course of events. » (Parfit 1987 : 23) Décrire la personne comme nous ou non ne serait alors qu'un choix à propos de la manière d'appliquer notre concept de personne.

Bref, l'expérience de pensée des spectres de connectivité montre que la question de notre identité numérique n'aurait d'importance et n'exprimerait une différence réelle que si nous étions des entités ayant une existence propre, indépendante de celle de leurs constituants, et dont la survie est une question « tout-ou-rien ». Mais puisque tout indique que le dualisme est faux, cela implique selon Parfit que nous ne sommes que des entités conceptuelles. Notre identité numérique se réduirait alors généralement à l'identité qualitative, c'est-à-dire à l'exacte similarité (R&P : 241-2). La seule exception qui fait en sorte que Parfit ne juge pas le concept d'identité numérique inutile réside dans les cas hypothétiques où les relations de continuité seraient dédoublées (*branching*) – si on est dupliqué par exemple – ces relations étant alors à elles seules insuffisantes pour distinguer l'un de l'autre deux individus exactement similaires. L'identité numérique ne serait pertinente que lorsque l'*unicité* d'un individu est contestée.

1.2.2 – L'animal et le réductionnisme constitutif

Dans *Reasons and Persons*, Parfit n'envisage que le dualisme cartésien comme version plausible de la *ego theory* et n'accorde aucun égard à la possibilité que nous soyons fondamentalement des animaux. Nous allons donc la considérer ici. L'animalisme peut-il être une formulation soutenable de la *ego theory* et expliquer pourquoi nous serions une entité ayant une existence propre et une identité déterminée ? Si c'est le cas, la conclusion de Parfit suite à son argument des spectres nous aurait alors conduit à un faux dilemme : l'échec du dualisme cartésien n'impliquerait pas la victoire du réductionnisme ontologique. Si nous ne sommes pas des âmes, cela ne veut pas nécessairement dire que la *bundle theory* est vraie à notre sujet, et ce même si elle l'était à propos des personnes.

Afin que l'animalisme se qualifie comme version plausible de la *ego theory*, il faudrait qu'un animal soit une entité réelle, c'est-à-dire dont l'existence ne se réduit pas à celle d'un corps, d'un cerveau, et d'un ensemble d'événements physiques et mentaux. Cela semble difficile à défendre et explique peut-être pourquoi Parfit ne considère pas l'animal comme candidat de la *ego theory*. Sommes-nous en mesure de justifier que les animaux échappent au réductionnisme ?

Commençons d'abord par évaluer les implications d'appliquer ce réductionnisme aux animaux. Cela signifierait que les animaux ne seraient pas des entités réelles, mais seulement des *bundles* conceptuels. Cette position semble tout aussi difficile à défendre que l'alternative : si les animaux ne sont pas des entités réelles, c'est-à-dire des objets matériels situés dans le monde, mais sont réductibles aux entités qui les constituent et n'existent alors qu'en vertu de notre façon de parler, dans notre schème conceptuel, qu'est-ce alors qu'un objet « réel » ?

Parfit qualifie ainsi l'existence indépendante d'autres entités :

« Entities are independent, or separately existing, not absolutely, but in their relation to other entities. In relation to the plinth on which it stands, a statue is a separately existing entity. Cartesian Egos, if they existed, would be not only distinct from human bodies, but also, in relation to these bodies, independent entities. » (Parfit 2007 : 5)

Puisque que l'existence indépendante est relative selon Parfit, sa position implique donc que les éléments d'un *bundle* pourraient très bien être eux-mêmes des *bundles* d'entités

plus élémentaires. Le réductionnisme constitutif pourrait donc s'étendre à tous les objets couramment considérés comme réels – même aux corps, aux cerveaux et aux expériences qui constituent notre existence en tant que personnes. Nous ne serions alors que des *bundles* de *bundles*. Ultimement par contre, le *bundle* le plus élémentaire devrait déboucher sur des entités réelles, les « vrais » objets dont est constitué le réel, ayant une existence indépendante de toute autre entité. Si on se fie à la science actuelle, ces entités seraient des particules microphysiques – quoique nous semblions capables de creuser toujours plus avant dans le domaine de l'infiniment petit et de découvrir que les entités qu'on considère élémentaires ne sont en fait que des *bundles* d'autres entités. Il serait toutefois absurde de défendre que *it's bundles all the way down*. Par contre, on pourrait tout autant avancer, comme l'a fait Hume en élaborant sa *bundle theory*, que le réel est plutôt constitué de perceptions – les particules microphysiques se réduisant elles-mêmes à leurs effets observables et/ou mesurables.

Certes, il est tout de même contre-intuitif d'ôter toute réalité aux objets que nous considérons habituellement comme réels, et d'affirmer que personnes, tables, roches, et chiens n'existent qu'en vertu de notre façon de parler parce qu'ils se réduisent aux choses dont ils sont constitués. Notre ontologie serait alors bien différente de celle adoptée par l'homme de la rue. Mais si c'est à cela que mène l'argument de Parfit, sa force serait sapée : nous révéler comme autre chose que ce qu'on croit intuitivement être, comme n'existant qu'à l'état de pauvre entité conceptuelle, ne serait pas davantage choquant si tout objet qu'on pense réel était en fait aussi un *bundle* conceptuel. L'application de la *bundle theory* à nous-mêmes ne serait pas un *further fact* par rapport à la *bundle theory* générale.

D'ailleurs, certains auteurs défendent l'idée que la *bundle theory* serait en fait une *explication* de ce qu'est un objet réel (Robinson 2014 ; Rettler et Bailey 2017), et qu'elle n'impliquerait donc pas le réductionnisme constitutif – malgré que nous les ayons utilisés comme synonymes jusqu'ici. Ce serait ce qui relie les éléments d'un *bundle* ensemble qui en ferait un objet réel. Il n'est cependant pas nécessaire d'évaluer la signification de la *bundle theory* ici ; il suffit de constater que la portée du réductionnisme constitutif pour notre conception de nous-mêmes a une ampleur bien moins grande que celle que lui confère Parfit.

À ce sujet, Peter van Inwagen fait dans *Material Beings* une analyse qui est éclairante. Il se préoccupe de la question suivante : sous quelles conditions est-ce qu'un ensemble d'entités compose proprement un objet matériel, c'est-à-dire réel ? Pour mieux comprendre le problème métaphysique, il est utile de distinguer les notions de constitution et de composition employées par les auteurs :

Constitution : Selon Parfit, lorsque des entités en *constituent* une autre, l'existence de l'entité constituée est réductible à celle de ses constituants. L'entité constituée est alors conceptuelle, alors que les constituants ont une existence indépendante et sont donc des entités réelles, à moins qu'ils soient eux aussi constitués d'autres entités.

Composition : Selon van Inwagen, lorsque des entités en *composent* une autre, l'entité composée a une existence réelle, en tant qu'objet matériel, et n'est donc pas ontologiquement réductible. Les composantes ne sont alors que des parties de l'entité réelle, et ce sont donc celles-ci qui n'ont pas d'existence propre, étant seulement des entités conceptuelles, des « objets virtuels » dans sa terminologie.

Ainsi, lorsqu'il y a constitution, il n'y a pas composition, et vice versa. Déterminer si l'animal peut effectivement être considéré comme une entité réelle revient alors à se demander s'il est *constitué* de divers d'éléments, ou plutôt *composé* de ceux-ci.

Nous n'allons cependant pas nous engager davantage dans une discussion sur la méréologie qui dépasse largement le cadre de ce mémoire, quoi qu'elle serait nécessaire pour répondre adéquatement à Parfit. Nous allons nous limiter à discuter d'arguments en faveur de l'idée que le réductionnisme constitutif ne pourrait s'appliquer aux animaux, *même s'il s'appliquait* aux personnes, aux tables et aux roches.

Locke nous met en garde contre l'idée que les êtres vivants ne seraient que des ensembles de constituants, car cela n'en ferait que des agglomérats de matière. Ce qui les différencie, c'est que leurs constituants sont *organisés* d'une manière particulière, cette organisation étant en quoi consiste la *vie* :

« We must therefore consider wherein an Oak differs from a Mass of Matter, and that seems to me to be in this; that the one is only the Cohesion of Particles of Matter any how united, the other such a disposition of them as constitutes the parts of an Oak; and *such an Organization of those parts, as is fit to receive and distribute nourishment* so as to continue and frame the Wood, Bark, and Leaves, *etc.*, of an Oak, *in which consists the vegetable Life.* » (Locke 1975 : II.XXVII, §4 – c'est moi qui souligne)

La *bundle theory* semblerait donc omettre ce principe d'organisation, sans lequel les

constituants ne se composent ni en un tout ni en une vie, pourrait-on dire en y voyant une affinité avec van Inwagen.

Parfit pourrait certes rétorquer que les relations causales entre les éléments du *bundle* qui assurent son unité et son identité rendent suffisamment comptent d'une telle organisation fonctionnelle. Cependant, constater qu'il y a un principe d'organisation à l'œuvre ainsi que des relations causales entre les éléments d'un *bundle*, entre nos événements et états mentaux dans notre cas, ne saurait expliquer *pourquoi* c'est le cas. Leibniz objecte d'ailleurs à Locke que l'organisation ne saurait être suffisante pour expliquer la persistance de l'identité numérique, parce qu'elle pourrait subsister alors même que l'individu devient numériquement autre :

L'organisation ou configuration, sans un principe de vie subsistant que j'appelle *Monade*, ne suffirait pas pour faire demeurer *idem numero*, ou le même individu, car la configuration peut demeurer individuellement. (Leibniz 1966 : II.XXVII, §4).

Leibniz donne l'exemple d'un fleuve dont l'eau change toujours, et celle du bateau de Thésée continuellement réparé ; ceux-ci conserveraient leur organisation fonctionnelle alors même qu'ils perdent leur identité numérique. L'existence continuée d'un être vivant requerrait donc, sans surprise, l'identité de la *vie*, qui donne à l'individu son identité numérique et *explique* son organisation fonctionnelle et donc les relations causales entre les éléments qui le composent. La vie ne se résumerait donc pas à une simple organisation de parties.

Or, si l'identité des êtres vivants consiste en l'identité de leur vie, il semble difficile de concevoir que leur existence soit ontologiquement réductible. En effet, la vie ne paraît pas être quelque chose de divisible en un nombre d'éléments constitutifs auxquels elle se réduirait, ni pouvoir tenir à un degré réduit comme la connectivité psychologique ou physique. Si je coupe mon doigt, ma vie n'est pas alors divisée en deux ; mon doigt est « mort » et ma vie demeure pleinement mienne. Ce doigt coupé, par contre, serait alors réductible aux particules élémentaires qui le constituent. Selon van Inwagen, il ne serait donc pas à proprement parler un objet réel, mais virtuel (van Inwagen 1990 : 170).

Parfit propose certes des expériences de pensée mettant en jeu la division des hémisphères cérébraux dont nous allons traiter plus loin (1.3), mais même si elles avaient

les implications qu'il leur prête, ce ne seraient pas des cas de division de la vie mais plutôt de son dédoublement. De plus, la science actuelle, malgré ses efforts, n'arrive pas à reproduire la vie à partir des éléments chimiques qui y prennent part. Il nous manquerait peut-être bien ce qui assure leur *composition* en un tout dynamique. Nous en sommes à croire qu'il faudrait, en plus de ces composantes, des conditions environnementales particulières extrêmement difficiles à reproduire. De récentes recherches de la NASA suggèrent même que certains éléments complexes dont la vie dépend, présents dans les molécules d'ARN et d'ADN, ne pourraient voir le jour que dans l'espace, et possiblement seulement lors de la formation de planètes ou de l'explosion d'étoiles (Marlaire 2015). Bref, malgré l'ignorance entourant le sujet, il semble que nous ayons des raisons de croire que le réductionnisme constitutif ne pourrait être applicable aux êtres vivants.

Selon van Inwagen, si le réductionnisme constitutif s'applique à presque tous les objets que nous considérons réels, il demeure tout de même deux types d'entités concrètes, véritablement réelles, dans notre ontologie : les particules élémentaires, et les êtres vivants.¹⁰

« [T]hings compose a thing if and only if their collective activity constitutes a *life*. A life is a special kind of self-maintaining event into which particles of matter are constantly being drawn, and from which, normally after a fairly short period of participating in the life, they are expelled. Thus, the only composite objects are living organisms. » (van Inwagen 1993 : 684)

Ce serait seulement dans le cas des êtres vivants que la composition de particules élémentaires formerait un tout ayant des pouvoirs causaux intrinsèques, un principe d'auto-organisation non réductible à celui de ses composantes, et dont on pourrait affirmer qu'il a une existence propre.

Dressons un bilan des possibilités quant à la réalité des êtres vivants. Une première possibilité serait que le réductionnisme constitutif soit applicable à tout objet considéré réel, et qu'il n'y ait donc jamais composition de particules élémentaires en objets matériels complexes. Si tel est le cas, la réduction ontologique de notre existence ne saurait nous embêter plus que la réduction d'une roche aux particules qui la constituent ; nous ne serions pas un cas spécial d'entité que nous croyons faussement être

¹⁰ Selon van Inwagen, les particules élémentaires ne se qualifient pas clairement comme des *objets matériels* à proprement parler, quoi qu'elles se comportent tout comme et partagent leurs caractéristiques principales (van Inwagen 1990 : 19).

réelle. Cela ne nous empêcherait pas d'affirmer que nous sommes des entités aussi réelles et matérielles que des chaises, rendant l'argument de Parfit stérile. Les problèmes de persistance d'entités complexes seraient certes éliminés puisque celles-ci n'existeraient pas ; cependant, ce serait aux tenants d'un tel réductionnisme constitutif universel de justifier l'idée que tous les objets matériels seraient en fait virtuels ou conceptuels, et qu'il n'existe réellement que des particules.

Ensuite, une seconde possibilité serait celle où tous les objets matériels couramment conçus comme réels le sont effectivement, où tous les *bundles* de particules se composent en objets matériels, et où le réductionnisme constitutif ne s'applique donc pas. Malgré les problèmes que pourrait avoir une telle position, notamment celui de justifier ce qui cause la composition dans chaque cas, l'irréalité des êtres vivants n'en serait pas un.

Finalement, il reste un ensemble de possibilités où le réductionnisme constitutif s'appliquerait à un sous-ensemble des objets considérés réels, alors que la composition vaudrait pour le reste. Cependant, si le réductionnisme constitutif s'appliquait aux animaux, quelles raisons aurions-nous alors pour justifier qu'il ne s'applique pas également aux autres objets considérés réels ? Quelles raisons aurait-on de croire que les vivants seraient ontologiquement réductibles mais pas les tables, par exemple ? S'il y a une raison plausible pour expliquer qu'une entité échappe au réductionnisme, la vie semblerait la meilleure candidate à notre disposition.

En somme, il semble plausible de défendre l'idée que les animaux ont une existence réelle, non-réductible à celle des éléments qui les composent, ou du moins qu'ils sont des entités pas moins réelles que tout autre objet matériel complexe. Si cette ébauche d'une position ne fait pas erreur, l'animal humain serait alors un candidat qualifiable pour la *ego theory*.

1.2.3 – L'animalisme et nos croyances à propos de notre identité

Examinons maintenant les implications de l'existence réelle des animaux humains sur leur identité. Lorsque nous avons présenté la *ego theory*, nous avons expliqué qu'un bon nombre de croyances intuitives à propos de notre identité lui sont associées. Énumérons-les de nouveau. D'abord, le fait que notre identité soit maintenue ou non,

c'est-à-dire que nous survivions ou non à un événement, attesterait d'une différence dans le réel plutôt que seulement dans notre manière de nous exprimer. La préservation de notre identité numérique serait d'ailleurs ce qui importe le plus pour nous lorsqu'on se soucie de notre survie. De plus, nous croyons que notre identité serait une question « tout-ou-rien », c'est-à-dire qu'elle ne pourrait tenir à un degré réduit, mais serait soit pleinement présente ou absente. Finalement, notre identité serait déterminée, c'est-à-dire qu'il y aurait toujours une réponse non-arbitraire à la question de savoir si un être humain est le même ou non. Or, afin que l'animalisme se qualifie comme version de la *ego theory*, les croyances qui en découlent devraient correspondre à celles que nous venons d'énoncer, ou alors il faudrait justifier pourquoi ça n'est pas le cas mais que l'animalisme est tout de même une version légitime de la *ego theory*. Voyons si cela est soutenable.

D'abord, nous avons vu que selon Parfit, si nous sommes des entités réelles, il serait raisonnable de croire que notre identité dans le temps est un fait qui n'est pas seulement conceptuel, réductible à des faits de continuité physique et/ou psychologique, mais qu'il met en œuvre un *further fact*. Ce fait additionnel, si nous sommes des animaux, ce serait la continuité ou non de notre vie – nous en avons discuté dans la sous-section précédente.

Il est toutefois possible de se demander si le fait que nous soyons des entités réelles *requerrait* l'irréductibilité de notre identité à des faits de continuité. La vie serait-elle réductible à des faits de continuité biologique ? Un animaliste qui n'est pas réductionniste à propos de notre existence pourrait-il raisonnablement adopter un critère réductionniste de l'identité ? On pourrait soulever l'argument qu'une pomme, qu'on considère être un objet matériel réel, semble avoir un critère d'identité physique, son identité pouvant être réduite à des faits de continuité spatio-temporelle. Un *further fact* ne serait pas nécessaire, et s'il en allait de même pour nous, un non-réductionnisme à propos de notre existence n'en impliquerait alors pas un à propos de notre identité. John McDowell, qui croit que nous sommes fondamentalement des animaux, défend une telle position (1997 : 245).

Cependant, est-il vraiment cohérent d'affirmer qu'alors même que l'existence d'une entité ne se réduit pas à celle d'un ensemble de constituants, son identité serait réductible à des faits à propos de ces mêmes constituants ? Cela serait plausible, si c'est la

composition de ces éléments qui en fait une entité réelle – quoique l'inverse soit également concevable. Si nous prenons en compte la discussion précédente sur la constitution et la composition, les pommes ne sont peut-être pas des objets réels ; en tant que système de dissémination de la semence du pommier, elles ne sont peut-être pas elles-mêmes des entités vivantes. C'est peut-être seulement pour cette raison que leur identité serait réductible. Il faudrait donc démêler les notions de constitution, de composition et de vie avant de s'avancer sur la réductibilité de notre identité. Par contre, un non-réductionniste à propos de notre existence pourrait tout de même faire usage d'un critère réductionniste d'identité, sans considérer que celui-ci fournit les conditions nécessaires et suffisantes pour le maintien de notre identité, mais seulement une justification raisonnable de le croire.

Cela dit, même si les faits de notre identité et de la continuité de notre vie étaient réductibles à des faits de continuité de l'organisme et n'étaient donc que conceptuels, il n'en reste pas moins que si nous sommes des animaux, ces faits seraient à propos d'une entité réelle plutôt que conceptuelle, et attesteraient donc d'une différence dans le réel plutôt que dans notre choix de vocabulaire. Si mon identité n'est pas préservée, cela signifie que l'animal que je suis est mort. Une description de moi vivant ne pourrait en être une description équivalente. Cela expliquerait notre croyance intuitive que l'affirmation de notre identité exprime un fait réel.

Par suite, comme seuls les faits exprimant des différences dans le réel sont importants selon Parfit, le fait que la persistance de notre identité serait synonyme du maintien de notre vie expliquerait en partie notre croyance qu'elle est ce qui importe le plus lorsqu'on se soucie de notre bien-être futur. Certes, si notre identité était réductible, l'importance qu'on lui accorde ne serait que dérivée des faits auxquels elle se réduit ; cependant, cette importance serait d'autant plus grande puisqu'à propos d'une entité réelle plutôt que conceptuelle. Néanmoins, Eric Olson, qui défend l'animalisme classique que nous allons voir au second chapitre (2.1), se range du côté de Parfit en admettant que notre identité n'est pas ce qui importe le plus dans notre souci de soi. La préservation de nos contenus mentaux, qui serait indépendante de notre identité animale, pourrait être plus importante que notre survie biologique. Il faut donc une raison supplémentaire à celle que nous sommes une entité réelle afin de soutenir que notre identité serait ce qui

importe le plus. Nous tenterons de justifier cette croyance, contre Olson, lorsque nous proposerons une version alternative de l'animalisme à la toute fin de ce mémoire (3.2.4).

Évaluons maintenant si, en conformité avec la *ego theory*, notre identité en tant qu'animaux serait déterminée et ne pourrait tenir à un degré partiel. Parfit, comme nous l'avons vu, croit que ces croyances vont de pair : c'est seulement si notre existence en tant qu'entité réelle était une affaire de tout ou rien plutôt que de degrés, qu'il serait alors raisonnable de croire que notre identité est déterminée. Or, une position animaliste semble avoir de telles implications, puisqu'elle rattache notre identité au maintien de notre vie animale. Comme nous l'avons suggéré dans la sous-section précédente, la vie ne semble pas être quelque chose qui peut n'être qu'en partie maintenu : qu'on me coupe un membre, et la vie le quitte mais reste pleinement mienne. Peu importe l'état de notre santé, la vie semble demeurer pleinement nôtre tant que nous ne sommes pas morts. Par conséquent, l'animalisme semblerait fournir à première vue une justification de la croyance selon laquelle notre identité est déterminée, et qu'il y aurait une démarcation claire dans les spectres de connectivité de Parfit, c'est-à-dire un moment où le retrait d'une cellule de trop ou dont la fonction est essentielle signerait notre arrêt de mort.

Cependant, la détermination de notre identité n'est peut-être pas aussi simple, même si nous sommes des animaux dotés d'une existence réelle. Van Inwagen défend que leur existence n'est pas indubitablement une affaire de tout ou rien. Si les conditions nécessaires et suffisantes pour que des éléments composent un être vivant sont vagues, alors notre existence en tant qu'objet réel pourrait être indéterminée. Y a-t-il un moment déterminé où une masse de cellules embryonnaires passe de l'état d'objet virtuel à celui d'objet complexe, réel et vivant ? Van Inwagen doute qu'il y ait une telle démarcation entre ces deux états et conclut qu'il doit donc y avoir des cas limites de composition et donc de vie (van Inwagen 1990 : 238). De tels cas seraient également possibles en fin de vie.

Or, s'il y a des cas limites d'*existence* des vivants, où un arrangement de particules se rapproche d'un cas de composition mais n'y est pas tout à fait, il y aurait alors selon van Inwagen des cas limites d'*identité* de la vie. Il propose d'imaginer un *spectre de perturbations de la composition*, allant de la continuité sans accroc de la composition de nos parties en un être vivant jusqu'à sa désintégration – si une bombe nous réduisait en

bouillie, par exemple. Comme pour les spectres que Parfit imagine, ce sont les cas du centre qui sont problématiques, où il n'est pas clair que la perturbation est temporaire et ne porte pas atteinte à l'intégrité de la vie (van Inwagen 1990 : 241). Si l'identité de l'animal dépend de l'identité de sa vie, et qu'il pourrait être indéterminé que nous ayons la *même* vie suite à une perturbation, plutôt qu'une *nouvelle* vie résultant d'une re-composition de nos parties en un tout vivant, notre identité pourrait alors être indéterminée. Que nous soyons des entités réelles ne garantirait pas que notre identité soit toujours déterminée.

Van Inwagen ne propose toutefois aucun exemple hypothétique de ce que serait une perturbation de la composition qui ne serait pas fatale, laissant au lecteur la tâche d'en imaginer. Il n'est pas du tout clair que ce qu'on pourrait imaginer serait effectivement une perturbation de la *composition*, plutôt que seulement une perturbation de la santé ou du comportement alors même que la composition est maintenue. Si on écarte cette possibilité, il ne resterait alors que l'indétermination à propos du commencement et de la fin de l'existence, desquelles on pourrait s'accommoder : une telle indétermination n'affecterait pas la détermination de l'identité *durant* l'existence de l'être vivant. La vie resterait une affaire de tout ou rien, même si les processus de composition et de décomposition de la vie étaient graduels. L'indétermination apparente proviendrait-elle alors uniquement d'un flou dans nos concepts de composition, de vie, et de partie d'un tout ?

La position dominante, quoique contestée, veut que toute indétermination de l'identité soit linguistique plutôt qu'ontologique : ce serait nos mots qui sont vagues, pas les choses. Locke est de cet avis :

« [T]he difficulty or obscurity that has been about this Matter, rather rises from the Names ill-used, than from any obscurity in things themselves. For whatever makes the specifick *Idea* to which the name is applied, if that *Idea* be steadily kept to, the distinction of anything into the same and divers will easily be conceived, and there can arise no doubt about it. » (Locke 1975 : II.XXVII, §28)

Si nos mots étaient suffisamment précis et non-ambigus, l'indétermination serait alors levée. Bertrand Russell défend également une position voulant que tout flou soit propriété de nos représentations seulement. Il ne le voit pas simplement comme un défaut, mais comme un atout d'une connaissance bourgeonnante, à préciser par la suite

pour augmenter sa valeur pragmatique : « a vague belief has a much better chance of being true than a precise one, because there are more possible facts that would verify it. » (Russell 1923 : 91)

Leibniz fournit un argument substantiel en faveur de l'idée que l'identité numérique des choses *doit* être ontologiquement déterminée. Si ce n'était pas le cas, il n'y aurait tout simplement pas d'objets distincts :

Si deux individus étaient parfaitement semblables et égaux, en un mot, *indistinguishables* par eux-mêmes, n'y aurait point de principe d'individuation ; et même j'ose dire qu'il n'y aurait point de distinction individuelle, ou de différents individus à cette condition. [...] il y en aurait alors d'indistinguishables en soi, et qui ne pourraient être discernés que par des déterminations extérieures, sans fondement interne, ce qui est contre les plus grands principes de la raison. Mais la vérité est que tout corps est altérable, et même altéré toujours actuellement, en sorte qu'il diffère en lui-même de tout autre. (Leibniz 1765 : II.XXVII, §3)

Si les objets réels n'avaient pas d'identité numérique absolument distincte de celle des autres objets, le réel serait organisé de façon contradictoire, deux objets « distincts » pouvant alors se confondre au point où ils ne seraient pas des objets différents. La possibilité même d'un monde où il y a des objets distincts requiert donc que chaque objet soit *unique*, c'est-à-dire que l'identité de tout objet soit ontologiquement déterminée, même si nous avons de la difficulté à la déterminer. Le principe d'identité des indiscernables formalisé par Leibniz, selon lequel tout objet qui partage toutes ses propriétés avec un autre objet est en fait le même objet, découle du rejet de la contradiction ontologique aberrante qu'entraînerait une indétermination de l'identité numérique. La même intuition est d'ailleurs résumée chez Quine par le slogan *no entity without identity* (Quine 1969 : 23). L'identité numérique des objets réels ne pourrait donc jamais être équivalente à l'identité qualitative.

Certes, pour des entités seulement conceptuelles, peut-être que le principe qui les distingue n'a pas à être interne, mais pourrait être externe, défini par nous, et donc épistémique plutôt qu'ontologique. L'affirmation de Parfit que l'identité numérique n'est que l'identité qualitative tiendrait peut-être pour de telles entités. Encore faudrait-il départager les entités réelles des entités conceptuelles. Et pour éviter les confusions, peut-être serait-il préférable d'abandonner le langage de l'identité numérique et ne parler que d'identité relative dans ces cas. Cependant, même si les chaises n'étaient que des *bundles*

conceptuels, il semble tout de même contentieux d'affirmer que le principe de distinction déterminant leur identité ne proviendrait que de nous. Leibniz affirme même que des ombres ont une identité numérique déterminée, à l'origine de la possibilité de les distinguer malgré qu'elles puissent se confondre en se superposant (Leibniz 1966 : II.XXVII, §1).

Gareth Evans s'appuie sur le principe d'identité des indiscernables de Leibniz pour défendre qu'il ne pourrait y avoir d'objets vagues, quoique son argument grandement discuté demeure contesté. Résumons-le. Considérons *a* et *b*, dont nous supposons qu'il est indéterminé s'ils sont identiques. Cela implique que *b* a la propriété d'avoir une identité indéterminée avec *a*. Or, il est faux que *a* a une identité indéterminée avec *a* ; *a* diffère donc de *b* sur au moins une propriété. Par le principe d'identité des indiscernables, puisqu'ils ne partagent pas toutes leurs propriétés, on doit alors conclure que *a* est de façon déterminée non-identique à *b*. Evans en conclut que l'indétermination ne peut alors se trouver dans le réel (Evans 1978).

Harold Noonan remarque que l'argument d'Evans ne saurait montrer qu'il ne peut y avoir d'*objets* vagues, seulement que même si un objet était vague, son identité serait tout de même déterminée :

« [A]ll it proves is that every vague object is determinately distinct from every precise object and every other vague object. So it is consistent to hold both that there are vague objects and that the identity relation is precise. »
(Noonan 2004 : 131)

Selon Noonan et bien d'autres, il est évident qu'il y a des objets *spatialement* vagues, comme les nuages et les montagnes dont les limites physiques sont incertaines, tout comme des objets *temporellement* vagues, comme les êtres vivants dont le commencement et la fin d'existence ne semblent pas clairement déterminés. S'il n'a pas tort, l'indétermination dans le processus de composition de la vie ne conduirait pas à l'indétermination de l'identité des êtres vivants, contrairement à ce qu'en dit van Inwagen. Il serait toutefois possible de contester l'existence *réelle* d'objets spatialement vagues, en affirmant que nuages, montagnes et autres entités similaires sont réductibles aux particules les constituant, et n'existent que comme entités conceptuelles.

Il y en aurait bien plus à dire sur la question de la détermination de notre identité ; nous allons nous limiter ici à un nombre restreint d'affirmations. Afin de savoir si notre

identité est indéterminée ou non, il faudrait répondre à la question de l'origine de l'indétermination : vient-elle exclusivement de notre langage et de notre compréhension imparfaite, ou y a-t-il une indétermination ontologique, à même la structure des choses et du réel ? De plus, la réponse à la question de l'origine de l'indétermination semble liée à la question de la composition, c'est-à-dire au problème de distinguer les entités réelles des entités conceptuelles. Si une entité n'est que conceptuelle, il semble justifiable de dire que son identité pourrait être indéterminée si notre concept est vague – comme Parfit le soutient au sujet des personnes – ou encore conventionnellement déterminée si le concept est précis ou arbitrairement précisé – comme il serait possible de le faire pour un tas de sable. Dans le cas d'une entité réelle, il serait justifiable de croire que son identité est déterminée si la théorie linguistique de l'indétermination n'a pas tort et que l'indétermination est symptôme de notre ignorance (quoique des arguments allant dans l'autre sens soient également avancés). On peut se demander par contre si les défenseurs d'une identité numérique *ontologiquement* indéterminée prennent l'indétermination propre à nos concepts et aux entités conceptuelles pour preuve d'une indétermination intrinsèque aux objets réels.

Compte tenu de ce qui précède, la croyance intuitive que notre identité est déterminée semble justifiable si nous croyons être des animaux, que les animaux sont des entités réelles, que la continuité de la vie est une question « tout-ou-rien », et que l'indétermination apparente provient de notre ignorance. La croyance que notre identité est déterminée dépend de plus de croyances que n'en laissait entendre Parfit, mais celles-ci paraissent toutes raisonnables. Elles ne seraient pas simplement fausses et à rejeter.

En somme, l'animalisme, moyennant une argumentation serrée, semble pouvoir rendre compte de nos croyances intuitives à propos de nous-mêmes, et pouvoir être compris comme une version de la *ego theory*. Notre identité numérique exprimerait un fait réel, à savoir l'existence d'un être vivant, et ne pourrait alors tenir à un degré réduit. Elle serait déterminée, hormis possiblement aux extrémités de notre carrière si la composition et la dé-composition étaient des processus graduels. Et elle pourrait être ce qui importe le plus lorsqu'on se soucie de soi, comme nous le défendrons vers la fin de ce mémoire (3.2.4).

D'ailleurs, si la *ego theory* exprime notre conception pré-théorique de notre

identité et un réalisme à propos de notre nature ontologique, ce n'est peut-être pas parce que nous sommes foncièrement dualistes, mais plutôt parce que nous nous concevons avant tout comme des animaux, c'est-à-dire comme des agents dans un monde physique. Après tout, nous ne saisissons les âmes que de manière négative, celles-ci n'étant descriptibles que comme ce qui est *immatériel*, *immortel*, et ainsi de suite, suggérant que c'est le monde naturel qui est notre domaine premier. Le domaine de ce qui n'est pas physique ne serait que le lieu où nous plaçons l'ensemble des choses que nous n'arrivons pas encore à comprendre pleinement comme faisant partie intégrante de la nature. De plus, les rituels entourant le corps après la mort – l'enterrement, l'embaumement et la momification pour préserver le corps – sembleraient indiquer que, même si l'on croyait que la vie après la mort était possible, il fallait du moins prévenir que cette « perturbation » de notre composition devienne permanente et provoque une décomposition dont nous ne pourrions revenir. Comme le suggère Locke, notre conception intuitive de l'être humain serait celle d'un animal d'une certaine forme, même si on croit qu'il a une âme :

« For I presume 'tis not the *Idea* of a thinking or rational Being alone, that makes the *Idea* of a *Man* in most People's Sense ; but of a Body so and so shaped joined to it [...] » (Locke 1975 : II.XXVII, §8)

Bref, nous, tout comme nos lointains ancêtres, sommes peut-être plus matérialistes qu'on ne le suppose en ce qui a trait à nos croyances à propos de notre identité et de notre nature ontologique.

1.3 – La division et l'unité de la conscience

Un deuxième argument majeur de Parfit en faveur de son réductionnisme et contre la *ego theory* s'appuie sur les données scientifiques à propos de la *division des hémisphères cérébraux*. Lorsqu'il s'agit de préciser notre nature en tant qu'êtres humains, outre les questions de notre persistance dans le temps, il faut également que notre réponse sache rendre compte de l'unité de notre vie mentale. La possibilité de la division du support de la conscience a des implications importantes à ce sujet, et par suite sur notre réponse quant à notre nature ontologique et notre existence. Nous allons commencer par exposer ici en quoi consiste le problème de l'unité de la conscience, avant de nous

attaquer au problème que pose la division et les conclusions que Parfit en tire (1.3.1). Nous verrons par la suite des recherches qui tendent à montrer que la division n'aurait pas les implications que Parfit lui accorde, et n'invaliderait pas la *ego theory* (1.3.2).

Notre conscience comporte deux unités à expliquer : l'unité à chaque moment (synchronique), et l'unité durant toute une vie (diachronique). D'une part, lorsque je croque dans un fruit par exemple, l'expérience du goût n'est pas une expérience qui a lieu séparément de celle de la texture, ni des pensées que j'ai au même moment ; nos expériences nous apparaissent comme un tout unifié. C'est ce que Parfit et d'autres nomment le caractère co-conscient de nos expériences. D'autre part, nos expériences passées, présentes et futures font partie d'une même vie mentale malgré leur séparation temporelle.

La *ego theory* et la *bundle theory* donnent différentes réponses au problème de l'unité de la conscience. Comme mentionné plus tôt, la *ego theory* y répond de façon simple et intuitive : nos expériences sont co-conscientes et forment *une* vie mentale parce qu'elles sont celles d'*une* entité, le sujet d'expérience (*nous*). En d'autres termes, l'unité du sujet et de sa vie expliquerait l'unité de ses expériences à tout moment et dans le temps. Selon Parfit, seul le fait que le sujet serait une entité ayant une existence réelle saurait justifier un tel fait brut (rappelons que Parfit identifie *sujet* et *personne*) :

« According to some Non-Reductionists, [...] [w]e must simply claim that these different thoughts and experiences are all had by the same person. [...] But that could not be true of persons, I believe, if, while distinct from their bodies, they are not separately existing entities. A person, so conceived, is not the kind of entity about which there could be such irreducible truths. »
(Parfit 2007 : 6)

Or, pour la *ego theory*, nous, le sujet d'expérience, avons une existence réelle. Selon certains d'ailleurs, la grande difficulté à expliquer l'unité de notre conscience sans en appeler à un sujet qui procéderait à cette unification *démontre* que le sujet a bien une existence réelle. Ce serait alors notre principe ontologique d'individuation qui, en nous engendrant comme entité concrète ayant des pouvoirs d'auto-organisation, serait également à l'origine de l'unification de nos expériences en une vie mentale. Par suite, expliquer en quoi nous sommes une entité réelle persistante telle une âme ou un animal expliquerait pourquoi notre conscience est unifiée.

Pour un réductionniste cette explication n'est pas disponible, parce que le sujet

n'est tout simplement pas une entité réelle et ne peut donc être le principe unificateur de notre vie mentale. Si on défend un critère d'identité physique, l'unité de la conscience pourrait alors être expliquée par le fait que toutes les expériences d'un sujet sont des changements d'état du même cerveau, ou y sont causalement reliées (Parfit 2007 : 6). Parfit s'oppose toutefois à une telle explication, parce qu'elle est l'expression d'un réductionnisme identifiant le sujet à son cerveau, alors qu'on s'en distingue dans notre usage, se considérant comme *ce* qui a un cerveau. Si on défend comme le fait Parfit un critère d'identité psychologique, alors l'unité de la conscience s'explique de la même manière que l'identité : les relations causales entre nos expériences, qui les relient en un *bundle* et assurent la persistance de notre identité, sont également ce qui unifie notre vie mentale dans le temps. Pour une telle position, notre identité *est* celle de notre conscience unifiée. La continuité psychologique et les chaînes de connectivité sous-jacentes font tout le travail explicatif.

Pour ce qui est de l'unité de notre conscience à chaque moment particulier, Parfit croit qu'une description de ce fait est une explication suffisante :

« This unity does not need a deep explanation. It is simply a fact that several experiences can be *co-conscious*, or be the objects of a single state of awareness. » (R&P : 250)

Notre conscience serait unifiée simplement parce que les multiples expériences qui ont lieu au même moment apparaissent dans un seul et même état de conscience. La circularité de cette explication n'attesterait que du fait que la co-conscience de nos expériences est un fait brut, n'appelant pas d'explication additionnelle. Cependant, cela contrevient à l'affirmation de Parfit que c'est seulement si le sujet était une entité réelle que l'unité de notre conscience ne requerrait pas plus de justification. Si cette explication de la co-conscience de nos expériences semble insatisfaisante à notre avis, elle demeure selon lui la seule à être plausible lorsqu'on considère l'argument de la division.

1.3.1 – L'argument de la division des hémisphères cérébraux

Entrons dans le vif du sujet et examinons l'argument de la division qu'élabore Parfit. Cet argument se base sur les recherches sur la division des hémisphères cérébraux effectuées par Roger Sperry et ses collègues dans les années 50-60, et pour lesquelles Sperry a gagné le prix Nobel de médecine en 1981. Leurs résultats ont été présentés à la

communauté philosophique par Nagel (1971), et sont d'ailleurs une des raisons ayant motivé Parfit à s'engager en philosophie (Parfit 1987 : 19). Les expérimentations de Sperry sur des chats et des chimpanzés ont conduit à la découverte que le corps calleux reliant les hémisphères cérébraux est le moyen principal de communication de l'information entre les hémisphères (Nagel 1971 : 399). Ce fait a alors été exploité pour soulager des patients sévèrement épileptiques dont les crises violentes voyageaient d'un hémisphère à l'autre par l'entremise du corps calleux. Leur corps calleux fut sectionné pour limiter la gravité de leurs crises et des séquelles, l'opération donnant lieu à ce qui est communément connu comme les cas de « cerveau divisé » (*split-brain*).

Sperry et Michael Gazzaniga – son étudiant à l'époque – se sont alors penchés sur ces cas pour voir l'impact de l'opération sur le comportement des patients, puisque chaque hémisphère ne contrôle qu'une moitié du corps et ne traite qu'une moitié des stimuli. Plus précisément, il y a une inversion structurelle : l'hémisphère droit contrôle les membres du côté gauche et traite l'information du champ visuel gauche, et inversement (Sperry 1984 : 661). Les résultats des chercheurs suggèrent que le fait de sectionner le moyen de communication entre les hémisphères a pour conséquence que chaque hémisphère « ignore » ce qui se passe dans l'autre et fonctionne de manière autonome. Lorsqu'un objet est présenté dans le champ visuel gauche d'un patient, et que les expérimentateurs contrôlent les données visuelles pour s'assurer que l'autre hémisphère n'y a pas également accès en bougeant les yeux, cet objet n'est alors reconnu que par l'hémisphère droit, ce qu'on constate au fait que le patient ne peut le signaler qu'avec sa main gauche, que cet hémisphère contrôle. Il en va inversement pour un objet présenté dans le champ visuel droit.

De plus, l'individu ne peut décrire verbalement que ce qu'il voit dans son champ visuel droit, puisque, chez la plupart des êtres humains, c'est l'hémisphère gauche qui seul traite le langage (Sperry 1968 : 731). Un patient pouvait donc dessiner de sa main gauche l'objet qu'il voyait dans son champ visuel gauche, tout en niant verbalement qu'il y avait là un tel objet parce que l'hémisphère gauche traitant le langage n'y avait pas accès. Le patient ignorait d'ailleurs pourquoi il l'avait dessiné lorsqu'on le lui demandait, croyant même que quelqu'un *d'autre* devait l'avoir fait (Sperry 1968 : 733). La division des hémisphères semblerait donc donner lieu à deux flux de conscience distincts chez le

même être humain, divisant l'esprit, selon les chercheurs :

« Each hemisphere can be shown to experience its own private sensations, percepts, thoughts and memories that are inaccessible to awareness in the other hemisphere. [...] In this respect each surgically disconnected hemisphere appears to have a mind of its own. » (Sperry 1984 : 663)

Malgré cela, les patients maintiennent un comportement normal dans leurs activités quotidiennes après une période de réadaptation, lorsque l'information visuelle n'est pas artificiellement ségréguée pour les besoins de l'expérimentation (Sperry 1968 : 724 ; Nagel 1971 : 401).

En partie en raison de ce comportement unifié, certains ont questionné le fait qu'il y aurait réellement chez ces patients deux flux de conscience, soutenant que seul l'hémisphère capable de langage recelait un esprit conscient, alors que l'hémisphère dit « mineur » avait des capacités approchant au plus celles d'autres animaux, et ne traitait l'information reçue qu'inconsciemment (Sperry 1984 : 665). Les recherches de Sperry réfutent ce modèle de la conscience, montrant que l'hémisphère mineur est capable de pensée délibérée et abstraite au-delà même des capacités des chimpanzés : il lui est possible, par exemple, de faire des opérations arithmétiques simples telles des additions de petits chiffres et d'associer des stimuli visuels à des objets différents qui « vont avec » (Sperry 1968 : 731). L'activité consciente ne serait donc pas limitée au seul hémisphère capable de discours.

Les expérimentations de Sperry et Gazzaniga montrent même que l'hémisphère mineur est capable de comprendre modérément bien des instructions verbales ou écrites, et peut associer mots et objets (Sperry 1968 : 731). Il peut même épeler et lire, par l'entremise de lettres présentées sur un support tactile, des mots de trois ou quatre lettres (Sperry 1984 : 665). Parfit souligne d'ailleurs que des individus qui subissent un accident vasculaire cérébral dans l'hémisphère dominant régressent parfois aux capacités linguistiques d'un enfant de trois ans, mais qu'ils peuvent développer de nouveau leurs capacités linguistiques avec l'hémisphère mineur (R&P : 246). S'il y a donc une spécialisation des hémisphères, chacun développant des capacités complémentaires, l'aptitude au langage articulé ne serait pas non plus rattachée à un hémisphère. Mais même si certaines capacités linguistiques l'étaient, nous pouvons imaginer que dans le futur un implant pourrait accorder ces fonctions à l'hémisphère mineur. Parfit considère

donc, pour son argument, que les hémisphères pourraient être fonctionnellement identiques, et produire deux flux de conscience pleinement développés et autonomes.

Parfit conclut de l'apparente divisibilité de la conscience que l'unité de celle-ci ne peut alors être expliquée comme le ferait un tenant de la *ego theory*, par le fait que les expériences sont celles d'un seul sujet d'expérience. Notons que l'argument de Parfit suppose que nous sommes essentiellement des personnes, et que le terme *sujet d'expérience* n'en est qu'un synonyme ; nous allons généraliser l'argument pour qu'il soit également applicable contre l'animalisme en plus du dualisme que vise Parfit. S'il y a effectivement deux flux de conscience chez les patients au cerveau divisé, l'unité de chacun des flux ne peut alors être expliquée par le fait que les expériences s'y trouvant sont possédées par le même sujet d'expérience, *nous* – fussions-nous une âme, un animal, ou une personne ayant une existence seulement conceptuelle – parce qu'une telle explication confondrait les deux flux en un. En attribuant les expériences dans les deux flux à une seule entité, on tenterait d'expliquer la présence de deux unités de conscience en les réduisant à une seule, et on ne rendrait donc pas compte du phénomène (R&P : 249). Simplement, comme il semble pouvoir y avoir deux consciences unifiées chez un même animal, l'unité de chacune serait inexplicable par l'attribution des expériences des deux consciences à cet animal – et pareillement pour une âme. L'explication que fournit la *ego theory* ignorerait les conséquences avérées de la division des hémisphères, et ne pourrait donc être correcte.

Il serait certes possible de répondre que suite à la division, nous serions en présence de deux sujets d'expérience occupant le même corps, chacun unifiant les expériences de leur flux de conscience respectif. L'explication de l'unité de la conscience proposée par la *ego theory* serait sauvée. Toutefois, cela impliquerait qu'il y aurait alors, selon notre position sur la nature des êtres humains, deux animaux, deux âmes, ou deux entités conceptuelles suite à la division. Cette solution n'est pas sans ses problèmes, en particulier en ce qui concerne notre identité numérique. Si nous « devenions » deux individus, lequel serions-nous, puisque l'identité est ici une relation à un seul terme ? Et comment justifierions-nous notre choix, si les deux consciences étaient psychologiquement identiques et en continuité avec nous d'avant l'opération – s'il est même justifiable d'affirmer que nous sommes l'un des deux esprits résultants ?

Outre ces problèmes d'identité, Parfit propose une expérience de pensée qui élimine la possibilité de *notre* division alors même que notre conscience se divise : la *division temporaire*. Il nous demande d'imaginer un individu ayant, en plus de deux hémisphères fonctionnellement identiques, un implant qui lui permettrait de bloquer la communication entre ceux-ci et de la rétablir à volonté (R&P : 246). Cela donnerait à un tel individu la capacité de traiter des problèmes ayant plusieurs solutions apparentes plus efficacement, divisant son esprit pour tester séparément deux différentes solutions dans ses deux flux de conscience résultants, pour, par la suite, les réunir et mettre leurs découvertes en commun. Parfit donne l'exemple d'un individu qui divise temporairement son esprit pour se donner toutes les chances de bien répondre à une question d'examen de physique pour lequel il ne lui reste que quelques minutes. Dans un tel cas, Parfit croit qu'on ne peut légitimement affirmer qu'il y a temporairement deux êtres humains :

« Given the brief and modest nature of this disunity, it is not plausible to claim that this case involves more than a single person. Are we to suppose that, during these ten minutes, I cease to exist, and two new people come into existence, each of whom then works out one of the calculations? » (R&P : 248)

Si nous croyons qu'il y a temporairement deux êtres humains, alors non seulement la division temporaire impliquerait-elle la mort de l'humain original et la création de deux nouveaux individus partageant sa mémoire, mais la réunion subséquente des deux flux de conscience signifierait que ces deux individus mourraient à leur tour, et que l'original reviendrait à la vie avec leurs mémoires combinées. Ça n'exigerait rien de moins qu'une succession de miracles. Par suite, si, lors d'une division temporaire de notre esprit, il ne subsiste qu'un être humain, quelle explication donner alors pour l'unité de chaque flux de conscience ?

Si un défenseur de la *ego theory* voulait conserver son explication de l'unité de la conscience, il se verrait alors forcé de distinguer entre les sujets d'expérience et l'être humain (R&P : 249). Il devrait affirmer que dans les cas de division temporaire, et par suite dans tous les cas de cerveau divisé s'il veut être cohérent, il y aurait chez un même être humain deux sujets correspondant aux deux flux et expliquant l'unité de chacun. Cependant, en ayant recours à une telle stratégie, la position d'un *ego theorist* perdrait sa plausibilité. Ce serait dire que nous, qui sommes sujet d'expérience en temps normal, ne

le serions plus si notre cerveau en venait à être divisé. Il y aurait alors dans une même vie à la fois des sujets qui seraient des êtres humains et des sujets qui ne le seraient pas (R&P : 250). Un patient au cerveau divisé était un sujet d'expérience avant l'opération, il ne le serait plus après. Quel genre de chose serait alors un sujet d'expérience ?

Si nous, êtres humains, sommes des entités réelles – âmes ou animaux – comme l'avancent les défenseurs de la *ego theory*, il serait absurde de croire que les sujets d'expérience seraient également de telles entités si elles en sont distinctes. Cela reviendrait à dire qu'un patient au cerveau divisé serait « devenu » en fait trois entités réelles suite à l'opération : une âme avec deux sous-âmes, ou un animal avec deux sous-animaux par exemple. Selon Parfit, lorsque nous supposons que le sujet d'expérience était l'être humain (c'est-à-dire la personne pour lui), il demeurait plausible de croire en son existence réelle ; mais maintenant que le sujet serait distinct de nous, il devient éminemment douteux de postuler qu'il y a de telles entités dont la seule raison d'être serait d'expliquer l'unité de notre conscience (R&P : 250). Le sujet ne serait alors qu'une invention pour colmater les trous dans notre connaissance, sans réellement expliquer le phénomène. Par contre, si les sujets d'expériences n'étaient pas des entités réelles, ils ne seraient alors pas le type d'entité qui pourrait plausiblement expliquer l'unité de la conscience puisque tout fait à propos d'elles se réduirait à des faits à propos de leurs constituants. La divisibilité de la conscience rendrait donc l'explication proposée par un *ego theorist* impossible à défendre, que les sujets d'expérience aient une existence réelle ou conceptuelle.

Pour les dualistes cartésiens, la situation est encore pire que pour les animalistes. S'ils veulent maintenir une explication de l'unité de la conscience conforme à la *ego theory*, cela impliquerait que la division des hémisphères cérébraux conduirait à la division de l'âme. L'âme ne serait alors pas si distincte du corps matériel, contrairement à ce qu'en croient les dualistes, puisqu'on affirmerait que son état et son unité dépendent de l'état du cerveau. Elle ne serait également pas non plus une entité simple et indivisible. Vouloir maintenir le dualisme pour expliquer la division reviendrait à nier ses préceptes fondamentaux.

En somme, si l'on s'en tient aux résultats de Sperry et Gazzaniga ainsi qu'aux expériences de pensée que Parfit a construites sur leur base, la *ego theory* fait face à de

sérieuses difficultés. Nagel, dans sa discussion de ces résultats et de leurs implications sur le nombre d'esprits qu'on attribuerait aux patients au cerveau divisé, énonce le problème ainsi :

« If I am right, and there is no whole number of individual minds that these patients can be said to have, then the attribution of conscious, significant mental activity does not require the existence of a single mental subject. »
(Nagel 1971 : 409)

Le sujet d'expérience, sans son pouvoir explicatif de l'unité de notre conscience, ne serait plus un concept épistémiquement pertinent. La division de la conscience semble donc, de par ses implications, indiquer que la position réductionniste à propos de nous-mêmes est la plus plausible ; et ce, même si « l'explication » de la co-conscience de nos expériences proposée par Parfit serait à revoir.

1.3.2 – L'argument de la division contesté

Attardons-nous maintenant à de plus récentes recherches sur des patients au cerveau divisé, qui révoquent en doute les résultats de Sperry et ses collègues et qui nous fourniraient une piste pour rétablir la plausibilité de la *ego theory* tout en faisant échec aux expériences de pensée de Parfit. Yair Pinto et son équipe de l'Université d'Amsterdam ont revisité la question de l'effet de la division des hémisphères sur l'unité de notre conscience en testant extensivement deux patients, et sont parvenus à des résultats divergeant de ceux de l'équipe de Sperry sur des points cruciaux. Alors qu'ils confirment que sans la communication assurée par le corps calleux, les stimuli visuels ne sont pas partagés entre les hémisphères, l'unité de la conscience serait tout de même maintenue. Par conséquent, l'explication de l'unité de la conscience proposée par la *ego theory*, à savoir que nous, sujets d'expérience, sommes la raison de cette unité, serait de nouveau défendable. Examinons ces résultats.

Comme nous l'avons vu, l'argument principal en faveur du modèle désormais canonique de la conscience divisée défendu par Sperry et Gazzaniga est la dépendance de la réponse du patient au champ visuel dans lequel un stimulus est présenté – le patient reconnaissant le stimulus soit avec la main gauche si dans le champ gauche, soit avec la main droite et verbalement si dans le champ droit. Si ce modèle de la conscience était juste et que la division des hémisphères donnait bien lieu à deux *agents* conscients

séparés et autonomes, ayant chacun accès à un champ visuel différent et contrôlant des membres différents, alors cette dépendance de la réponse au champ visuel devrait toujours être observable. Toute exception mettrait le modèle en doute, indiquant que ces foyers de conscience ne seraient peut-être pas séparés, et donc qu'il n'y aurait pas deux agents mais bien un seul.

Or, les recherches de Pinto et son équipe montrent qu'il y a des exceptions voire une absence complète de dépendance de la réponse au champ dans lequel se trouve le stimulus visuel. Les patients étudiés ont été capables d'indiquer leur reconnaissance d'un stimulus présenté dans un champ visuel même à l'aide de la main contrôlée par l'hémisphère qui n'aurait pas accès à ce champ, ainsi que verbalement :

« Across a wide variety of tasks, we observed that split-brain patients could reliably indicate presence, location, orientation and identity of stimuli throughout the entire visual field regardless of how they responded. » (Pinto 2017b : 1232)

Ces patients ont également pu localiser avec précision l'emplacement du stimulus dans une proportion dépassant largement celle obtenue s'ils avaient simplement tenté de deviner (Pinto 2017b : 1233). De plus, leur performance dans ces tests était de niveau similaire indépendamment du fait qu'ils répondaient avec les moyens sous le contrôle de l'hémisphère ayant accès au stimulus ou non.

Devant ces étonnants résultats, les chercheurs se sont demandés si l'information visuelle était partagée entre les hémisphères malgré l'ablation du corps calleux. Ce dernier est certes le moyen de communication principal entre les hémisphères, mais il n'est pas le seul lien entre eux, quoique ceux qui restent permettent au plus un flot d'information de 1 bit par seconde, ce qui est moins que ce qui est véhiculé dans une conversation normale, sans parler d'informations visuelles (Pinto 2017b : 1236 ; Vrije Universiteit 2017). Il serait toutefois possible que les cerveaux des patients aient compensé pour la perte du corps calleux, puisqu'ils ont subi l'opération au moins dix ans avant les tests. Cependant, les chercheurs ont confirmé que l'information visuelle traitée dans un hémisphère n'était pas accessible à l'autre, comme dans les cas répertoriés par Sperry, puisque les patients étaient incapables de comparer des stimuli perçus dans un champ visuel avec ceux perçus dans l'autre (Pinto 2017b : 1233-4).

Pour s'assurer que les patients étaient bien conscients de ce qui se passait dans

l'autre hémisphère malgré qu'ils n'avaient pas accès à l'information visuelle qui s'y trouve, Pinto et son équipe ont demandé aux patients de donner des cotes de confiance à leurs réponses quant à la détection et la localisation des stimuli visuels. Leur performance était meilleure dans les tests où ils s'accordaient une haute cote de confiance, montrant que leurs réponses correctes étaient le résultat d'une activité consciente et intentionnelle (Pinto 2017b : 1235). Ces résultats indiquent que dans les cas de cerveau divisé, malgré que les informations visuelles sont ségréguées à l'hémisphère dans lequel elles sont traitées, l'unité de la conscience serait tout de même préservée, et ce sans nécessiter la communication massive entre les hémisphères assurée par le corps calleux (Pinto 2017c : 846).

Qu'est-ce qui expliquerait alors la grande différence entre les résultats de Sperry et ceux de Pinto ? Selon Pinto et son équipe, c'est la méthodologie qui serait en cause. Même les tests de Sperry et Gazzaniga présentaient des exceptions quant à la dépendance de la réponse sur le champ visuel. Sperry le reconnaît :

« Although the general picture has continued to hold up in the main as described, [...] striking modifications and even outright exceptions can be found among the small group of patients examined to date. » (Sperry 1968 : 733)

Ces exceptions n'ont cependant pas été prises en compte par les chercheurs lors de l'élaboration de leur modèle de la conscience divisée, malgré qu'elles pouvaient être significatives étant donné le nombre restreint de patients testés. Pinto note d'ailleurs que d'autres chercheurs ont également obtenu des résultats mettant en doute la dépendance du type de réponse au champ visuel (Pinto 2017c : 838).

De plus, les tests de Sperry demandaient aux patients de rapporter ce qui se passait exclusivement sous forme qualitative plutôt que quantitative (Pinto 2017c : 839), de sorte que des biais pouvaient facilement s'y introduire. Pinto rapporte qu'avec l'approche de son équipe, lorsque les consignes données aux patients étaient clarifiées pour éliminer biais et réponses vagues, que les réponses étaient quantifiées, et qu'ils ajoutaient des tests de contrôle, les réponses tendant à montrer qu'il y avait une dépendance sur le champ visuel disparaissaient complètement (Pinto 2017c : 839).

En somme, les recherches de Pinto et son équipe montrent que les différences dans le fonctionnement des patients au cerveau divisé ne permettent pas de conclure que

l'unité de la conscience est divisée et produit deux sujets lors de la division des hémisphères, en opposition au modèle de Sperry et Gazzaniga. Ces différences seraient d'ailleurs mieux expliquées par un modèle aux implications moins extrêmes, à savoir que chaque patient serait un seul sujet dont la perception visuelle n'est pas intégrée entre les hémisphères (Pinto 2017c : 845). Cela rendrait compte de l'inhabileté de ces patients à comparer l'information visuelle entre leurs hémisphères, tout comme de leur comportement unifié ainsi que de leur sensation d'être normal et en contrôle de tout leur corps (Pinto 2017c : 844). Pinto décrit donc la condition des patients au cerveau divisé ainsi : « 'split phenomenality' combined with 'unity of consciousness' » (Pinto 2017b : 1236).

Un autre type de cas nous fournit des données allant dans le même sens. Krista et Tatiana Hogan sont de jeunes filles jumelles siamoises reliées par leurs cerveaux, et présentent un trait unique : les thalamus dans leurs tronc cérébraux sont reliés par un « pont thalamique », qui assure des fonctions similaires au corps calleux, c'est-à-dire la communication entre leurs cerveaux et le transfert des informations sensorielles (Pyke 2017). Les jumelles partagent leur sens tactile, leur goût, et leur vision (quoique Krista n'a accès qu'à un seul œil de Tatiana). Selon leurs dires, elles sont même capables de communiquer leurs pensées sans recours à la voix, et leur mère atteste que leurs humeurs sont également transmises entre elles (Pyke 2017). Cependant, elles ont chacune une personnalité différente, des intentions, des goûts et des désirs distincts. Par exemple, Krista aime le ketchup et le poulet ; sa sœur Tatiana en partage le goût lorsque Krista en mange, mais Tatiana n'aime pas ça du tout, ce qui cause des conflits (Dominus 2011). De plus, l'une peut être endormie alors que l'autre est réveillée (Pyke 2017).

Nous avons là l'inverse d'un cas de cerveau divisé : les jumelles présentent une unité de leur perception, mais leur consciences sont distinctes. Les analyses de Pinto et le cas des sœurs Hogan sembleraient donc montrer que l'unité de la conscience ne dépend pas de l'unité ou non de la perception : il est possible d'avoir une conscience unifiée alors même que la perception est divisée, tout comme des consciences peuvent être distinctes malgré qu'elles partagent leurs perceptions. Un sujet ne serait alors pas simplement un *bundle* de perceptions, contrairement à ce qu'en disait Hume.

Que pouvons-nous conclure de plus de ces récentes découvertes ? D'abord, les

hémisphères seraient vraisemblablement insuffisants pour assurer l'unité de la conscience. Si, lorsqu'ils sont divisés, l'unité entre eux est tout de même maintenue, il y a de bonnes chances que ce qui produit cette unité n'est pas localisé dans les seuls hémisphères. Par suite, les hémisphères ne seraient pas les porteurs de notre identité numérique – pas à eux seuls, du moins. Serait-ce plutôt le tronc cérébral, dont la structure n'est pas intrinsèquement divisible comme le cortex cérébral ? Cela irait dans le sens de la position animaliste d'Eric Olson, que nous verrons plus loin (2.1).

Ensuite, ces résultats sembleraient indiquer que l'unité de la conscience dans le temps ne pourrait être expliquée seulement par des relations psychologiques entre les divers événements mentaux, contrairement à ce que défend Parfit. En effet, si l'unité est maintenue alors même que la perception est divisée, certaines expériences perceptuelles feraient partie de la même vie mentale malgré qu'il n'y a pas de relations psychologiques les reliant en une série. Les expériences perceptives de mon hémisphère gauche ne seraient pas psychologiquement rattachées aux expériences de mon hémisphère droit, de sorte que les relations psychologiques seules ne rendraient compte que de deux séries d'expériences en parallèle, plutôt que de leur unité. L'explication de l'unité de la conscience semblerait donc requérir une dimension physique ou biologique. Si nous ne faisons pas erreur ici, l'explication proposée par la *ego theory*, à savoir que c'est un sujet réel persistant dans le temps qui unifie nos expériences, serait alors dans sa version animaliste plus plausible que l'explication réductionniste. Simplement, nos expériences seraient unifiées parce qu'elles sont celles du même animal : *nous*. Cela ne voudrait toutefois pas dire que l'explication s'arrêterait là, mais que toute explication complémentaire de nature biologique, physique et/ou psychologique prendrait ce point pour fondement.

De plus, si la division des hémisphères ne divise pas la conscience, alors les autres expériences de pensée sur la division élaborées par Parfit, ainsi que les conclusions qu'il en tire, seraient invalidées. Notamment, Parfit propose d'imaginer un cas de double transplantation des hémisphères, chacun allant dans le crâne d'un jumeaux de l'individu original (R&P : 254). Puisque Parfit suppose que chaque hémisphère maintient sa propre conscience, cela résulterait en deux agents conscients et autonomes, en continuité psychologique avec l'original. De ce cas de « double survie », il conclut que notre unicité

et donc l'identité numérique importent peu, puisqu'on ne peut dire si l'original est l'un ou l'autre, et que tout ce qui importe, la continuité et la connectivité psychologiques, seraient présentes chez chacun (R&P : 261). Sans l'appui de tels cas de dédoublement, sa position réductionniste serait grandement affaiblie.

Sans division, il n'y aurait pas non plus de cas de fusion, où deux hémisphères sont reconnectés ensemble pour fusionner deux sujets en un seul. Nous aurions au plus des cas similaires à celui des jumelles Hogan, à savoir deux sujets distincts pouvant partager une partie de leurs expériences. Or, la fusion était l'argument principal de Parfit pour défendre que ce qui importe était une question de degrés : « Just as division shows that what matters in survival need not take a one-one form, fusion shows that it can have degrees. » (R&P : 298) Elle montrait que la connectivité, dans le cas de conflits de croyances, d'intentions, et d'autres connections psychologiques, ne serait alors maintenue que partiellement lors de la fusion, certaines connections devant être rejetées parce qu'incompatibles avec d'autres. Si la division et la fusion sont éliminées des possibilités, il y a un argument possible voulant que, si l'on défend un critère d'identité psychologique, seule la continuité importe plutôt que la continuité *et* la connectivité. Par suite, comme la continuité n'est pas une question de degré mais de tout-ou-rien comme l'identité numérique, et comme elle ne pourrait être dédoublée par division, il serait alors possible de défendre que c'est bien l'identité numérique qui importe.

Cela dit, considérons tout de même la possibilité que Pinto et son équipe se soient trompés et que la division des hémisphères donne réellement lieu à la formation de deux consciences distinctes en continuité avec l'originale. Cela impliquerait-il que notre identité ne serait pas celle de l'animal ? que nous ne serions que des entités conceptuelles ? et que l'identité numérique n'importerait pas en raison du dédoublement possible ? Rappelons que la division, c'est-à-dire l'auto-réplication, est la première forme de reproduction chez les êtres vivants depuis les êtres unicellulaires. Soulignons également que les êtres conscients sont en mesure d'engendrer d'autres consciences, qu'ils produisent dans leur organisme avant de les enfanter. La venue au monde d'une seconde entité consciente dans un même animal par processus de division n'est pas si farfelue que ça lorsqu'on la considère comme un phénomène de reproduction. D'ailleurs, dans le cas des êtres qui se reproduisent effectivement par division tels les amibes, considérons-nous

qu'ils n'ont pas d'identité numérique et qu'ils ne sont alors que des entités conceptuelles ?

Si nous croyons que les animaux sont des entités réelles et ont donc une identité numérique qui leur est conférée par leur principe ontologique d'individuation, il serait incohérent de faire une exception pour ceux qui se reproduisent par division. Par conséquent, à moins qu'on soit de l'avis que le réductionnisme constitutif s'applique universellement et qu'il n'y a réellement que des particules, il y aurait alors une manière de déterminer l'identité des amibes, et celle-ci serait alors applicable aux animaux humains s'ils pouvaient se diviser. Bref, même si la division des hémisphères donnait lieu à deux consciences distinctes, il serait encore justifiable, si nous croyons être des animaux, d'affirmer que nous sommes des entités réelles dont l'identité numérique est ontologiquement déterminée, et donc que l'unité de notre conscience est explicable par le fait nous, animaux humains, unifions nos expériences en une vie mentale.

Résumons. Parfit défend que l'unité de la conscience ne serait explicable que de manière réductionniste, parce que la division des hémisphères produirait deux consciences distinctes ayant chacune leur unité et dont on ne pourrait rendre compte en les attribuant toutes deux à un seul sujet d'expérience. Il faudrait donc faire appel aux connexions psychologiques dans chaque flux afin de les distinguer. Ce serait alors une raison additionnelle de rejeter la *ego theory*. Par contre, comme la division des hémisphères cérébraux ne semble en fin de compte pas affecter l'unité de la conscience du sujet, la raison d'écarter l'explication que propose la *ego theory* ne serait plus valable. Plus encore, une explication purement psychologique de cette unité serait erronée, car elle ne saurait expliquer que dans le cas d'un patient au cerveau divisé, il y a unité malgré l'absence de connexions psychologiques reliant les deux foyers de perceptions. L'explication que propose la version animaliste de la *ego theory*, à savoir que la conscience serait unifiée parce que les expériences seraient possédées par un même sujet réel qui les unifie, l'animal humain, pourrait s'avérer la meilleure à notre disposition.

1.4 – Le réductionnisme constitutif et la réification des expériences

Dans cette section, nous allons glisser un mot à propos d'une conséquence implicite au réductionnisme constitutif de Parfit tel que nous l'avons présenté, qui est

évoquée par plusieurs comme objection majeure à sa position, même par des auteurs réductionnistes et/ou ayant adopté une approche psychologique de l'identité personnelle. En effet, puisque notre existence conceptuelle se réduit selon Parfit à l'existence réelle d'un ensemble d'éléments, parmi lesquels seuls nos événements mentaux seraient essentiels, cela semble impliquer que ces derniers seraient des entités réelles, ayant une existence propre et indépendante de toute autre entité. Le réductionnisme de Parfit ne nierait donc notre existence comme sujet réel qu'au prix d'une réification de nos expériences. Une position selon laquelle nous sommes essentiellement des animaux humains plutôt qu'un *bundle* conceptuel d'expériences qui peuvent exister par elles-mêmes serait alors d'autant plus plausible. Parfit conteste toutefois cette lecture de sa position, affirmant qu'il voulait seulement montrer que nos expériences ont une *intelligibilité* indépendante d'un sujet d'expérience plutôt qu'une *existence* qui pourrait en être indépendante. Il admet cependant que la formulation de ses propositions ait pu induire ses lecteurs en erreur (Parfit 1999 : 221, 270n56). Nous nous attarderons à ses arguments après avoir examiné plus avant l'objection qui est soulevée contre lui.

Sydney Shoemaker est le premier à formuler cette critique, remarquant que Parfit se rapproche dangereusement de Hume en comprenant le sujet comme *bundle* d'expériences (Hume le concevait comme *bundle* de perceptions). Puisque Parfit compare, comme Hume, le sujet à la nation, cela implique que, tout comme les citoyens qui constituent une nation peuvent exister sans en faire partie, les expériences pourraient exister sans dépendre d'un sujet (1985 : 446). Hume croyait d'ailleurs que c'était le cas pour les perceptions :

« [S]ince all our perceptions are different from each other, and from every thing else in the universe, they are also distinct and separable, and may be consider'd as separately existent, and may exist separately, and have no need of any thing else to support their existence. » (Hume 2007 : 1.4.5, 153 ; SB 233)

John Campbell (1992 : 386), Quassim Cassam (1992 : 362) et Simon Blackburn (1997 : 183) reconnaissent tous explicitement que la position initiale de Parfit implique que les événements mentaux seraient des entités ontologiquement indépendantes de sujets.

Shoemaker énonce ainsi son désaccord avec cette idée :

« This suggests that [Parfit] thinks that persons are logical constructions out of entities whose existence does not require that they be states of persons or

other 'subjects'—entities whose existence is not 'adjectival' on mental subjects, in the way seeings are adjectival on seers, deeds are adjectival on doers, or (for a non-mental example) dents are adjectival on dentable surfaces. » (Shoemaker 1985 : 446)

Selon Shoemaker, ainsi que Blackburn qui reprend également l'idée d'une *dépendance adjectivale*, tout comme des égratignures ne pourraient exister ni même être concevables sans surface égratignable, la nature même des expériences implique qu'elles ont une dépendance conceptuelle et ontologique sur des sujets d'expérience, celles-ci étant avant tout des « *experiencings* » (Shoemaker 1985 : 446). Blackburn remarque que c'est là une objection proprement kantienne, selon laquelle l'existence d'un sujet serait la condition de possibilité même d'expériences et de perceptions (Blackburn 1997 : 183-4).

John McDowell formule une critique qui va dans le même sens, mais qui insiste plutôt sur l'idée que le réductionnisme de Parfit implique que la conscience elle-même serait ontologiquement indépendante d'un sujet réel. En limitant le critère de la continuité de la conscience à la continuité d'éléments et de relations à l'intérieur de cette même conscience, le réductionnisme de Parfit élimine d'emblée la possibilité qu'elle soit le propre d'un sujet réel qui serait plus qu'une pure conscience, à savoir un animal humain (McDowell 1997 : 232-3). McDowell y perçoit la répétition de ce qu'il nomme « l'erreur cartésienne » (1997 : 235) :

« [T]hat utterly Cartesian conception of 'consciousness' as self-contained goes unquestioned in the implicit defense of reductionism that I have been considering. Indeed, we can see a Cartesian structure in Parfit's reductionism itself. » (McDowell 1997 : 233-4)

Le réductionnisme de Parfit ne différencierait significativement d'un dualisme que par son rejet de l'existence des âmes ; ce qui équivaut, pour qui conçoit la conscience comme unité autonome, au rejet de l'existence de tout sujet conscient réel.

Par contre, tenter de saisir le flux de conscience indépendamment de son appartenance à la carrière d'un sujet rendrait la conscience et les expériences inintelligibles selon McDowell ; Shoemaker, Blackburn, Campbell et Cassam sont également de cet avis, quoiqu'ils diffèrent partiellement quant à leurs raisons. Selon McDowell, un *quasi*-souvenir – c'est-à-dire un souvenir qui ne présuppose pas l'identité du sujet qui s'en rappelle – n'est compréhensible que de façon dérivée, en référence au concept ordinaire de souvenir qui, *lui*, implique l'identité de son sujet (McDowell 1997 :

241). L'intériorité ou la subjectivité de nos expériences ne pourrait non plus être comprise après coup, à partir de *quasi*-expériences indépendantes d'un sujet qui seraient présentées à la première personne à un sujet fictif construit de ces mêmes *quasi*-expériences (McDowell 1997 : 244). McDowell conclut alors que « the right basis for opposing reductionism is that we have been given no good reason to believe in a substratum of identity-free relations. » (McDowell 1997 : 245) Il n'y aurait pas de support conceptuellement plus simple auquel rattacher la conscience, que celui dont dépend l'identité du sujet réel.

Parfit suggère que la méprise quant à sa position, à savoir qu'elle impliquerait que les expériences peuvent exister sans sujet, proviendrait en bonne partie de la raison pour laquelle il dit éviter le concept *d'état mental* et lui préférer celui d'*événement* ou d'*expérience* pour décrire les éléments composant notre vie mentale (Parfit 1999 : 270n56). Il explique son choix ainsi dans *Reasons and Persons* : « it avoids what I believe to be one misleading implication of the words 'mental state'. While a state must be a state *of* some entity, this is not true of an event. » (R&P : 211) Par contre, si Shoemaker réfère à ce passage pour supporter son interprétation, cela n'explique pas qu'une bonne partie des commentateurs ait compris Parfit ainsi sans s'y référer. D'ailleurs, la présentation que nous avons donné de sa position dans ce chapitre n'en fait pas mention et s'appuie sur maints autres éléments, en particulier sur ses textes subséquents où il cherche à clarifier sa position.

Néanmoins, Parfit déploie un effort considérable pour répondre aux objections et défendre que l'intelligibilité des expériences indépendamment de sujets n'implique pas nécessairement la possibilité de leur existence indépendante. Son argumentation peut être résumée ainsi. L'idée d'une dépendance adjectivale invoquée par ses critiques, c'est celle d'une dépendance conceptuelle qui refléterait une dépendance ontologique. Si les égratignures dépendent conceptuellement de surfaces égratignables, c'est parce qu'elles en dépendent avant tout ontologiquement. Selon Parfit par contre, il n'en va pas ainsi pour les expériences et les sujets. Les expériences dépendent en effet conceptuellement de sujets *dans notre schème conceptuel*, mais s'il était possible de montrer que dans un *autre* schème conceptuel imaginable cette dépendance n'est pas nécessaire, alors celle qu'on constate dans notre schème serait ontologiquement insignifiante, plutôt

qu'indicative que les expériences dépendent ontologiquement de sujets (Parfit 1999 : 227). Si cela pourrait alors justifier que les expériences soient intelligibles indépendamment de sujets, ça n'implique pas pour autant qu'elles puissent exister par elles-mêmes, puisqu'elles pourraient quand même dépendre d'un support physique telles les cellules qui les réalisent. Examinons ces arguments plus en détail.

D'abord, Parfit réitère qu'il a toujours défendu que les expériences dépendent *conceptuellement* de sujets, même lorsqu'il affirmait qu'il serait possible de donner une description complète du réel qui omette de mentionner ces derniers :

« It would not be an open question whether these thoughts were had by a thinker, since the existence of this thinker was, I assumed, conceptually guaranteed. That is why I claimed that, in a full description, this fact would not need to be separately mentioned. » (Parfit 1999 : 252)

Il admet par contre que son rejet du terme d'*état mental* a confondu les choses : « It would have been clearer to say that, while all mental states must be, grammatically, states of some entity, that leaves it open what kind of entity this is. » (Parfit 1999 : 270n56) Ainsi, la dépendance conceptuelle des expériences sur un sujet n'impliquerait pas toutefois qu'il est autre chose qu'un sujet *grammatical*. Ce que Parfit a toujours refusé, c'est que ce sujet aurait une existence propre, réelle, et que les expériences en seraient *ontologiquement* dépendantes.

Là où Parfit admet s'être trompé, c'est dans son idée que notre schème conceptuel admettrait la possibilité qu'on se décrive pleinement sans référer à nous-mêmes ni même mentionner que nous existons. C'est un argument de Cassam qui l'a convaincu, malgré que Campbell défende le même, et que Thomas Nagel l'ait développé auparavant (1986 : chap. 4). En effet, sans supposer que les sujets existent, les énoncés auto-référentiels ne seraient pas objectivement compréhensibles. La valeur de vérité d'un énoncé tel « j'ai deux ans » dépend du sujet qui le dit, comme l'explique Cassam :

« Just as the content of a demonstrative thought is a function of the object referred to, so the truth condition and hence the content of a first-person thought is a function of the person referred to by the subject-term. In order to know *what* the thought is one must know *who* the thinker is. » (Cassam 1992 : 365)

Par conséquent, dans notre schème conceptuel, selon lequel les pensées sont le propre de penseurs, on ne peut se passer d'affirmer l'existence de sujets pensants. C'est pourquoi,

pour montrer que les expériences peuvent être intelligibles sans référence à des sujets qui en sont possesseurs, Parfit entreprend d'imaginer un schème conceptuel où celles-ci n'en dépendent pas, mais en sont plutôt conceptuellement séparables.

Mais avant d'entrer dans une telle élaboration, Parfit commence par expliquer en quoi le fait qu'une entité dépende conceptuellement d'une autre ne signifierait pas nécessairement qu'elle en dépende ontologiquement. Il donne un exemple qu'il tient de Peter Strawson (Parfit 1999 : 225-7). Selon Strawson, notre concept de *particule* dépend conceptuellement de notre concept d'*objet* ordinaire. En effet, alors qu'on peut avoir le concept d'*objet* sans celui de *particule*, l'inverse n'est pas vrai : nous ne comprenons ce qu'est une particule que comme une version microscopique d'un objet ordinaire, qui se comporte de façon similaire dans notre schème spatio-temporel. Mais malgré la dépendance conceptuelle des particules sur les objets, ce sont plutôt les objets ordinaires qui dépendent ontologiquement de particules. Parfit nomme cette dépendance ontologique *compositionnelle*, pour exprimer l'idée qu'un tout dépend des parties dont il est composé. Il est donc possible que les dépendances conceptuelle et ontologique tirent dans des directions opposées.

Or, c'est une dépendance compositionnelle que mettrait en jeu son réductionnisme ontologique du sujet. Un sujet se réduirait aux parties qui le composent, notamment à ses expériences, parce qu'il en dépend pour exister. S'il y a une dépendance adjectivale qui tient dans l'autre sens, à savoir des expériences sur les sujets, cela n'implique pas qu'il en va de même pour la dépendance ontologique. Par conséquent, la dépendance adjectivale des expériences sur les sujets pourrait très bien n'être que conceptuelle : « If experiences depend on subjects in some more important sense, we need more than grammar to show us that. » (Parfit 1999 : 238) Les relations entre concepts ne correspondent pas toujours à la réalité.

Notons toutefois que le langage de Parfit a changé : alors que son réductionnisme est partout formulé dans le vocabulaire de la *constitution*, il en appelle ici, et seulement ici, à la *composition*. Nous avons opposé ces concepts plus tôt en présentant l'ontologie de van Inwagen (1.2.2), et on peut se demander si Parfit omet de les distinguer ici. Alors que la constitution serait une relation qui réduit ontologiquement l'entité constituée à ses constituants, la composition réduit plutôt les parties au tout qu'elles composent. On

pourrait alors comprendre, contrairement à ce qu'en dit Parfit, que dans le cas d'une dépendance compositionnelle, ce seraient plutôt les parties qui dépendent ontologiquement du tout, parce que sans lui elles ne seraient pas des *parties* jouant un rôle fonctionnel dans un ensemble. Si cela était valable pour les sujets, les expériences dépendraient alors à la fois conceptuellement et ontologiquement des sujets qu'ils composent. Ainsi, l'explication de Parfit n'est pas si bénigne qu'elle n'en paraît et vient accompagnée de présupposés à propos de la méréologie qui sont peut-être faux.

Poursuivons néanmoins. Après avoir dissocié la dépendance adjectivale de la dépendance ontologique, Parfit avance que dans un autre schème conceptuel, les expériences pourraient conceptuellement dépendre d'autre chose que d'un sujet. Il imagine des êtres qui ne s'auto-attribuent pas leurs expériences et qui n'ont pas de concept équivalent à celui de *sujet* :

« In place of our concept of a person, they have concepts of two closely related entities: living bodies and unified sequences of interrelated mental processes and events, such as thoughts, experiences, and acts. » (Parfit 1999 : 227)

Dans leur manière de comprendre le réel, les expériences seraient conceptuellement dépendantes de corps et/ou de séries d'expériences plutôt que de sujets.

Un tel schème ne permettrait certes pas à ces individus de se concevoir comme possesseurs d'expériences, mais ils pourraient très bien comprendre l'intériorité des expériences selon Parfit, étant en mesure de les saisir comme ayant lieu *dans* un corps ou dans une série d'expériences, contrairement à ce qu'en dit McDowell (Parfit 1999 : 228). De plus, ces individus pourraient saisir l'aspect subjectif ou qualitatif des expériences, sans devoir référer à un sujet pour qui cet aspect est présent – si l'on en croit Parfit. Il explique que lorsqu'on compare nos expériences à celles de quelqu'un d'autre, nous n'aurions pas besoin d'imaginer comment les expériences de l'autre apparaîtraient *pour nous*, mais simplement comment elles apparaissent, indépendamment d'un sujet (Parfit 1999 : 257). Finalement, il ne serait pas nécessaire dans un tel schème de supposer l'existence de sujets pour comprendre le contenu de certaines pensées, puisque ces dernières ne seraient jamais exprimées à l'aide d'indexicaux auto-référentiels, mais utiliseraient plutôt des démonstratifs – « *cette* sensation dans *cette* série d'expériences » plutôt que « *ma* sensation », par exemple (Parfit 1999 : 230). La critique de Cassam et

compagnie serait contournée. Parfit en conclut alors que les expériences pourraient être pleinement intelligibles indépendamment de sujets, quoique cela requerrait que nous nous détachions de notre schème conceptuel usuel.

Parfit admet certes qu'il n'a pas montré que sa position écarte définitivement la possibilité de l'existence d'expériences isolées. Il croit par contre que le simple fait de rattacher chaque expérience à une série d'expériences particulière pourrait suffire à éliminer l'implication que les expériences pourraient exister par elles-mêmes, sans qu'on doive invoquer de dépendance adjectivale à un sujet (Parfit 1999 : 242). Cependant, il y a des arguments plus forts que celui de la dépendance adjectivale qui tendent à montrer qu'il ne pourrait y avoir d'expériences sans sujets, notamment le fonctionnalisme du mental de Shoemaker. Selon celui-ci, la nature même des événements mentaux est de jouer un rôle causal particulier dans un système mental auquel ils appartiennent – ce système étant justement ce que Parfit conçoit comme un *sujet* (Shoemaker 1985 : 446). Cette manière de concevoir les expériences s'apparente à l'idée que nous avons mentionnée plus haut, à savoir celle de parties qui se composent pour former un tout interdépendant. Si une telle position était correcte, les expériences dépendraient ontologiquement de sujets, contrairement à ce qu'en croit Parfit.

Parfit avance toutefois que même si la position fonctionnaliste de Shoemaker était juste et avait des implications ontologiques, celles-ci ne consisteraient pas nécessairement en l'attribution d'expériences à des sujets réels. Pour le défendre, il s'appuie de nouveau sur la possibilité de la division du sujet suite à la division des hémisphères du cerveau, que Yair Pinto a fortement remise en question comme nous l'avons vu plus tôt (1.3.2). Examinons tout de même l'argument.

La division du sujet implique qu'une même expérience pourrait appartenir à plusieurs sujets différents. En effet, suite à la division, les deux sujets résultants ne seraient pas l'original, mais auraient toutefois ses souvenirs. L'attribution d'expériences à un sujet particulier, l'animal avant la division par exemple, ne serait alors pas métaphysiquement pertinente (Parfit 1999 : 254). Il nous faudrait tout de même pouvoir identifier les expériences indépendamment du sujet réel. Parfit suggère donc que celles-ci seraient plutôt ontologiquement dépendantes non pas d'un sujet, mais de la partie du sujet qui les réalise : « Experiences could owe their identity, not to the person who has them,

but to something smaller: the brain, or group of brain cells, in which they are realized. » (Parfit 1999 : 248) Si Pinto a raison, les raisons que Parfit fournit ici pour rejeter la dépendance ontologique des expériences à un sujet seraient inappropriées. Par contre, même si d'autres raisons pourraient être avancées à cet effet, une position selon laquelle les expériences doivent être attribuées aux parties du sujet réel plutôt qu'au tout est lourde de conséquences ontologiques. Nous en traiterons au troisième chapitre (3.1), en réaction à la position finale de Parfit qui propose justement que nous ne serions qu'une partie de l'animal humain (2.3).

En somme, la position de Parfit, selon laquelle le sujet est un *bundle* d'expériences, est structurée comme celle de Hume et emploie les mêmes analogies. Il paraît donc légitime de croire qu'elle a les mêmes implications, à savoir que les expériences pourraient exister par elles-mêmes. Cependant, après une laborieuse entreprise de clarification, sa position peut plutôt se résumer ainsi : 1) nous sommes les sujets grammaticaux de nos expériences, et ce uniquement en vertu de notre schème conceptuel sans lequel nous n'existerions pas ; 2) ces sujets dépendent ontologiquement des expériences qui les constituent et s'y réduisent ; 3) les expériences dépendent à leur tour ontologiquement de leurs réalisateurs physiques, c'est-à-dire d'ensembles de cellules du cerveau. Si l'exposé de sa position dans *Reasons and Persons* omet cette troisième partie et évite de dire que le sujet se réduit directement aux cellules qui réalisent ses expériences, c'est peut-être bien parce que ces expériences pourraient être physiquement réalisées par d'autres ensembles de cellules voire par une autre sorte de support physique, et ne sont pas en elles-mêmes essentielles à notre identité. Enfin, en avançant que nos expériences dépendent ontologiquement de cellules, ainsi que conceptuellement de corps ou de séries d'expériences – dans un schème conceptuel alternatif certes – Parfit parviendrait à les rendre intelligibles sans références à un sujet – si et seulement si ses arguments tiennent. Mais comme nous le verrons au troisième chapitre, la dépendance de notre vie mentale à un sujet n'est pas si facilement évacuable.

*

Passons en revue certains points principaux de ce chapitre. Selon Parfit, nous n'avons pas d'existence réelle ; nous ne sommes que le *bundle* conceptuel réunissant les éléments qui entrent dans notre constitution, parmi lesquels seuls les événements

mentaux seraient essentiels. Notre identité se réduirait alors à des faits à propos de ces événements, et ce serait ceux-ci qui importent lorsqu'on se soucie de notre survie. Parfit propose deux arguments principaux en faveur du réductionnisme et contre la *ego theory* rivale. D'une part, l'expérience de pensée des spectres de continuité montre selon lui que la *ego theory* ne serait plausible que si nous étions des entités réelles distinctes de nos constituants, et dont l'existence serait une affaire de tout-ou-rien. Selon Parfit, seul le dualisme correspondrait à ces exigences, et les données indiquent qu'il serait faux. D'autre part, les cas de division des hémisphères cérébraux montreraient que seul le réductionnisme serait en mesure de rendre compte des données quant à l'unité de la conscience dans ces cas.

En réponse, nous avons défendu que l'animalisme pouvait être une version de la *ego theory* qui saurait respecter les contraintes que Parfit met en place ainsi qu'expliquer l'unité de notre conscience, puisque la division du cerveau ne produirait pas une division de la conscience. D'ailleurs, adopter l'animalisme n'impliquerait pas de rejeter le réductionnisme à propos des personnes, puisque nous pourrions alors avancer que nous sommes des animaux humains ayant une existence réelle, alors même que les personnes s'y réduisent et ne seraient en fait qu'une manière de saisir ces animaux. Cela dit, il convient de garder à l'esprit que les arguments invoqués de part et d'autre quant à notre nature ontologique s'inscrivent dans un débat plus large concernant ce qu'est un objet réel par opposition à un objet ontologiquement réductible, et dont l'influence va être visible dans l'entièreté de ce mémoire.

CHAPITRE II – L'ÉVOLUTION DE LA POSITION DE PARFIT

Dans le premier chapitre, nous avons exposé le réductionnisme que Parfit applique à l'existence des personnes c'est-à-dire à nous-mêmes, de même que les arguments principaux qu'il avance en faveur de sa position. Nous avons également présenté quelques objections qui représentent un premier pas pour contester l'idée que la sorte d'entité fondamentale que nous sommes n'est pas la *personne*, comme le défendent les lockéens tel Parfit, mais bien l'*animal*. Dans ce deuxième chapitre, nous allons voir plus en détail l'animalisme ainsi que les arguments principaux développés pour le défendre, arguments qui se sont imposés dans le débat et qui ont forcé les penseurs lockéens à adapter leurs conceptions (2.1).

Nous allons ensuite suivre l'évolution de la pensée de Parfit en réaction à ces arguments. Parfit se voit d'abord contraint de « battre en retraite » (Parfit 2012 : 9) et d'abandonner la distinction lockéenne entre animal et personne afin de conserver son approche psychologique de l'identité tout en affirmant que nous sommes des animaux (2.2). C'est seulement bien des années plus tard qu'il parviendra à formuler une position satisfaisante à ses yeux qui sache mieux répondre aux arguments animalistes tout en maintenant la distinction lockéenne : nous, les personnes, serions alors une partie de l'animal humain (2.3).

2.1 – La réplique animaliste

Dans cette section, nous allons nous concentrer sur la position animaliste. Nous allons d'abord examiner les critiques que les animalistes invoquent contre l'approche psychologique lockéenne, à savoir l'argument dit « du fœtus » (2.1.1) ainsi que celui dit « de l'animal pensant » (*The Thinking Animal Argument*) (2.1.2). Nous verrons par la suite l'animalisme lui-même, c'est-à-dire la thèse selon laquelle nous sommes numériquement identiques à un animal humain, et ce qu'elle implique quant à notre identité, la nature des personnes et l'identité personnelle (2.1.3).

2.1.1 – L'argument du fœtus et l'importance de notre vie mentale

Eric Olson développe l'argument du fœtus en faveur de l'approche biologique de l'identité dans son ouvrage *The Human Animal* (1997 : chap. 4). Son but est de montrer que si nous adoptons une approche psychologique de l'identité et que nous nous concevons alors comme des entités réductibles à leur vie mentale, cela implique de rejeter l'idée que nous sommes des animaux humains, parce que ceux-ci commencent leur existence sans vie psychologique. Nous examinerons ensuite certaines tentatives de réponse à cet argument.

Avant de rentrer dans le détail de l'argument, il convient de rappeler l'argument principal de l'approche psychologique rivale, commun à Locke, Parfit et Shoemaker, et contre lequel Olson élabore le sien. L'argument s'appuie sur un principe omniprésent chez Locke dans sa discussion sur l'identité :

« That therefore that had one beginning is the same thing, and that which had a different beginning in time and place from that, is not the same but divers. » (Locke 1975 : II.XXVII, §1)

David Wiggins le reformule ainsi :

Principe de la carrière : Si *a* et *b* sont [numériquement] identiques, alors ils doivent avoir la même carrière (*life history*). (Wiggins 2001 : 37 – c'est moi qui traduit)

Par suite, la stratégie pour défendre la thèse que la personne est une entité numériquement distincte de l'animal consiste à défendre l'idée que leurs carrières pourraient diverger, ce que font les partisans de l'approche psychologique en imaginant une multitude d'expériences de pensée très élaborées. Celles de Locke, que nous avons mentionnées en introduction, sont bien connues.

L'expérience la plus problématique pour l'idée que nous serions des animaux est celle de la transplantation de nos hémisphères cérébraux dans un autre organisme humain préalablement décervelé. Notre tronc cérébral serait laissé intact, de sorte à maintenir et réguler les fonctions vitales de notre organisme original. Notre vie mentale, quant à elle, serait transférée à l'organisme qui reçoit nos hémisphères. Selon les définitions lockéennes d'animal et de personne, l'un comme organisation fonctionnelle de parties propre à maintenir la vie (Locke 1975 : II.XXVII, §5-6), l'autre comme entité rationnelle capable de conscience de soi (Locke 1975 : II.XXVII, §9), l'animal et la personne que

nous étions avant l'opération survivraient de façon séparée. Les recherches de Roger Sperry évoquées dans le premier chapitre, qui indiquaient qu'un seul hémisphère cérébral serait suffisant pour maintenir la conscience, ont servi d'appui à cette conclusion.¹¹ Intuitivement, nous croyons que suite à l'opération, nous irions là où iraient nos hémisphères parce qu'ils servent de support à notre vie mentale, survivant ainsi en tant que la personne plutôt que l'animal original. Cette *Transplant Intuition*, telle que la nomme Olson, est invoquée comme argument décisif en faveur de l'approche psychologique de l'identité. Animaux et personnes seraient donc des entités distinctes ayant des conditions de persistance différentes, biologiques pour les uns et psychologiques pour les autres, parce que leurs carrières pourraient diverger.

Contre cette intuition, Olson oppose celle que toute personne humaine a déjà été un fœtus (Olson 1997 : 73). Cette intuition a également l'avantage d'être cautionnée par nos connaissances en biologie. Nous commençons à exister en tant qu'animal humain lorsque les cellules embryonnaires se différencient et s'harmonisent pour former un organisme multicellulaire plutôt qu'une simple agglomération de cellules identiques, c'est-à-dire environ quatorze à dix-sept jours après la conception (Olson 1997 : 91) – même si le moment exact de ce commencement ne serait pas déterminé si van Inwagen a raison (voir 1.2.3). C'est à ce moment que séparer la masse de cellules en deux ne résulterait pas en deux embryons plus petits, mais causerait plutôt la mort de l'organisme (Olson 1997 : 91).

Le problème pour l'approche psychologique, c'est qu'à ce stade l'embryon humain n'a même pas les rudiments d'hémisphères cérébraux nécessaires pour supporter des traits psychologiques. Olson rapporte que les hémisphères ne seraient pas fonctionnels avant le cinquième mois suivant la fertilisation, le fœtus ne faisant l'expérience d'aucun contenu mental – sans parler de comportement intentionnel, de représentation symbolique ni de raisonnement ; ceux-ci n'apparaissant qu'au cours des deux premières années suivant la naissance (Olson 1997 : 73-5). Un animal humain commence ainsi son existence sans vie mentale et peut la terminer pareillement, dans un état végétatif, de sorte qu'il n'y a pas continuité psychologique durant toute son existence. Avoir des propriétés mentales n'est donc pas nécessaire à la persistance des animaux humains, et la continuité psychologique

¹¹ Faisons pour le moment abstraction des recherches récentes de Yair Pinto qui réfutent ces résultats – nous y reviendrons dans le troisième chapitre.

ne peut être leur critère d'identité.

Par suite, l'approche psychologique implique au moins que les personnes humaines n'ont jamais été numériquement identiques à de jeunes fœtus, et qu'elles meurent lorsque « leur » animal perd ses capacités mentales et/ou son contenu mental, en tombant dans le coma par exemple. Plus encore, si on se fie à la définition lockéenne de la personne, qui requiert rationalité et conscience de soi, un partisan d'une telle approche devrait défendre l'idée que les personnes humaines n'ont jamais été de jeunes bébés, mais qu'elles sont venues au monde seulement lorsque ces bébés ont acquis de telles propriétés mentales. Bref, les prémisses qui rendent la *Transplant Intuition* plausible viennent accompagnées d'implications contre-intuitives.

Si l'on accorde plutôt plus de poids à l'intuition que nous avons commencé notre existence en tant que fœtus, l'expérience de pensée de la transplantation des hémisphères nous conduit à une conclusion différente, à savoir que nous sommes numériquement identiques à l'organisme comateux qui survit à l'opération, plutôt qu'à un des ses organes ou à sa vie mentale. Par conséquent, même si l'individu recevant nos hémisphères croirait être nous – puisqu'il aurait nos souvenirs – sa croyance serait tout simplement fausse. Olson ne nie pas toutefois que ce qui importe pour nous pourrait très bien être notre vie mentale plutôt que notre identité numérique, en accord avec Parfit (Olson 1997 : 56) ; mais malgré que cela expliquerait la *Transplant Intuition*, cela ne la rendrait pas juste pour autant. L'importance qu'a pour nous notre vie mentale n'implique pas que nous sommes numériquement identiques à cette vie mentale ni qu'elle est essentielle à notre persistance. Aux yeux d'Olson, l'expérience de pensée ne montre que le fait que si une telle opération était possible, un être humain pourrait alors être en continuité psychologique avec un individu autre que lui-même (Olson 1997 : 18 ; 2003 : 324).

Une réponse possible à cet argument, qui semblerait permettre aux partisans de l'approche psychologique d'affirmer que nous avons été un fœtus, consiste à dire que le fœtus est *devenu* la personne lorsque les propriétés mentales requises sont apparues. Mais une telle affirmation est hautement problématique si on maintient que l'animal et la personne sont numériquement distincts. La position néo-lockéenne implique que nous n'aurions pas là un cas analogue à la chenille se métamorphosant en papillon, c'est-à-dire à un processus où une entité demeure numériquement la même alors qu'elle subit un

changement qualitatif radical ; mais plutôt un cas où une nouvelle entité vient au monde lorsque la personne, c'est-à-dire *nous*, débute son existence (Olson 1997 : 79). Olson pose la question suivante : que se passerait-il alors avec le fœtus lorsque je viens au monde ?

Il serait bien commode pour les néo-lockéens que le fœtus disparaisse et soit remplacé par la personne qui prendrait alors contrôle du corps, remarque Parfit (2012 : 7), mais c'est là une possibilité franchement absurde :

« According to [it], it is logically impossible for a human fetus to come to be a normal, adult human being : the fetus *necessarily* ceases to exist as soon as its nervous system has developed enough to make thought possible. » (Olson 1997 : 79)

Les fœtus humains auraient donc une propriété unique dans tout le monde animal : celle d'avoir un code génétique qu'ils ne pourraient jamais réaliser avant qu'il ne cause leur mort. Gagner des capacités mentales supérieures serait alors une avenue évolutive stérile, provoquant l'extinction massive de toute espèce naturelle qui s'y engage. Mais ce n'est de toute évidence pas le cas, puisque les animaux humains continuent de se reproduire et d'engendrer des fœtus. Cette idée que la personne se substituerait à l'animal apparaît tout simplement incohérente. D'ailleurs, si la personne remplaçant le fœtus perdait ses capacités mentales durant sa vie, en tombant dans le coma ou en se faisant enlever ses hémisphères cérébraux par exemple, cela impliquerait que l'animal réapparaît ou renaît dans un état physiologiquement plus développé que lorsqu'il est disparu.

Pour expliquer la relation de la personne au fœtus, et le fait que nous sommes des êtres matériels, il ne reste donc apparemment que la possibilité suivante pour les néo-lockéens : que le fœtus survive à l'apparition de capacités mentales et poursuive son développement jusqu'à l'âge adulte, mais que la matière qui le compose en vienne à être partagée avec une deuxième entité qui en est numériquement distincte et possède ces propriétés mentales, la personne. Nous, les personnes, serions alors des animaux au sens où nous serions chacun constitués d'un animal, c'est-à-dire qu'un animal et personne coïncideraient dans le même espace. L'animal humain ne deviendrait jamais à proprement parler une personne, mais seulement son réceptacle, étant donné la possibilité que leurs carrières divergent. Shoemaker défend un tel *constitutionnalisme* :

« There is a sense in which each of us is something that was once a fetus—but this is an "is" meaning "is constituted by," not the "is" of identity. There is no need to say that the fetus went out of existence when the person came into

being; it continued to exist, eventually becoming a mature human animal. But at no point are the person and the human animal one and the same. » (Shoemaker 1999a : 498)

Cette manière de concevoir ce que nous sommes comporte de sérieux problèmes. Olson trouve d'abord que cette façon de dire que nous « sommes » des animaux tout en pouvant survivre sans continuer à l'être est hautement trompeuse :

« [T]hese philosophers are saying that each of us is a non-animal that relates in some intimate way to an animal. They put it by saying that we *are* animals because that sounds more plausible. This is salesman's hype, and we shouldn't be fooled. » (Olson 2003a : 319)

Si nous pouvons survivre en tant qu'autre chose que des animaux, nous ne pouvons pas leur être numériquement identiques. Le constitutionnalisme est même compatible avec l'idée que nous serions des substances immatérielles, la notion de constitution pouvant être utilisée pour affirmer que les âmes « sont » des animaux, et donc des objets matériels. Ainsi, le constitutionnalisme prétendrait fournir une réponse à l'argument du fœtus alors qu'il ne ferait que masquer les vestiges d'un dualisme cartésien.

Bref, l'argument du fœtus montre que si l'on veut affirmer que nous sommes des êtres matériels et que nous avons été des fœtus, un critère biologique de l'identité serait plus approprié qu'un critère psychologique. En réponse, le constitutionnalisme cherche à réconcilier tant bien que mal l'approche psychologique lockéenne avec l'idée que nous avons été des fœtus, mais comme on va le voir à l'instant, l'argument de l'animal pensant mine fortement sa plausibilité.

2.1.2 – L'argument de l'animal pensant et le constitutionnalisme

Voyons maintenant l'argument de l'animal pensant, qui s'appuie sur l'argument du fœtus et qui vise avant tout à contrer le constitutionnalisme. Certains éléments de cet argument ont été initialement présentés par Paul Snowdon (1990) et John McDowell (1997), qu'Olson a repris et considérablement développés. Son but est de montrer qu'une position néo-lockéenne selon laquelle la personne et l'animal seraient deux sortes d'entités fondamentales, à la fois numériquement distinctes mais coexistant dans le même espace, a des implications ontologiques inacceptables. Nous allons d'abord examiner ces conséquences, pour ensuite évaluer si le réductionnisme constitutif de Parfit – qui n'est

qu'une version du constitutionnalisme – saurait se soustraire à l'argument puisqu'il considère que la personne n'a pas d'existence propre en dehors de ses constituants. Nous passerons finalement en revue les manières de répondre à l'argument.

Supposons que conformément aux positions constitutionnalistes, personne et animal coïncident chez un même être humain et partagent leur matière, étant constitués des mêmes atomes sous la même configuration. Sous bien des formulations du monisme, les deux entités partageraient alors toutes leurs propriétés. Un physicalisme minimal, capable d'accommoder à la fois des positions éliminativistes, réductionnistes et non-réductionnistes, a notamment de telles implications. Il est communément construit sur la notion de survenance, c'est-à-dire sur l'idée que tout événement ou propriété non-physique – s'il y en a – dépend de et covarie avec des propriétés physiques (Kim 1998 : 14-5). Cela signifie qu'il ne pourrait y avoir de différence au niveau psychologique sans différence au niveau physique. Personne et animal seraient donc en tout point qualitativement identiques, puisque physiquement identiques. La seule différence entre eux se trouverait dans leurs conditions de persistance, tel que le postulent les lockéens.

Par suite, les mots que je, la personne, tape sur mon clavier sont également tapés par l'animal dont je suis constitué. Si je pense que je suis assis sur une chaise, l'animal le pense aussi. Nous utilisons tous les deux le même système nerveux, le même cerveau, et activons les mêmes neurones. Nos mots sortent de la même bouche. Olson conclut qu'il doit donc y avoir deux entités rationnelles et pensantes habitant notre corps et fonctionnant en parfaite synchronie :

« There are two thinking beings wherever we thought there was just one.
There are two philosophers, you and an animal, sitting there and reading this.
You are never truly alone: wherever you go, a watchful human animal goes
with you. » (Olson 2003a : 329)

C'est ce qu'Olson nomme le *Overcrowding Problem* et auquel Parfit réfère sous le nom du *Too Many Thinkers Problem*. En affirmant que la personne humaine est une entité constituée d'un animal humain mais numériquement distincte de lui, nous doublons le nombre d'entités rationnelles et pensantes dans notre ontologie tout comme le nombre de vies vécues – ce qui est pour le moins problématique.

Ce n'est là que le début des maux pour le constitutionnalisme néo-lockéen. Si l'animal et la personne pensent et ont les mêmes pensées, comment savoir lequel des

deux je suis ? Lorsque la personne et l'animal se disent « je suis une personne », l'un d'entre eux doit bien avoir tort, s'ils sont effectivement des entités numériquement distinctes. Cependant, puisqu'ils partagent leurs propriétés qualitatives, ils ont tous deux les mêmes raisons de croire qu'ils disent vrai. Nous pourrions donc très bien croire être la personne alors même qu'on se trompe :

« If you *were* the animal and not the person, you'd still think you were the person. So for all you know, you're the one making the mistake. Even if you are a person and not an animal, you could never have any reason to believe that you are. » (Olson 2003a : 330)

C'est ce qu'Olson nomme le *problème épistémique*. Si l'animal est distinct de la personne mais qu'il partage ses pensées, le référent du « je » deviendrait irrémédiablement ambigu, et nous n'aurions aucune manière non-arbitraire de déterminer la sorte d'entité que nous sommes ; et c'est sans parler de nos conditions de persistance.

Un troisième problème remet en question la raison même de distinguer numériquement la personne de l'animal. Snowdon remarque que si l'animal humain est rationnel, capable de pensée et de conscience de soi en vertu du fait qu'il coïncide avec la personne et partage ses propriétés, il se qualifie alors lui-même comme personne selon la définition lockéenne (Snowdon 1990 : 94). Par conséquent, non seulement le constitutionnalisme implique-t-il qu'il y a deux penseurs là où l'on croit qu'il n'y en a qu'un, mais également que ceux-ci seraient tous deux des personnes. C'est ce que Parfit nomme le *Too Many Persons Problem*. Chaque être humain serait en fait à la fois un « animal-personne » dont les conditions de persistance sont biologiques, et une « pure-personne » dont les conditions de persistance sont psychologiques, le premier survivant avec l'organisme alors que le second survivrait avec les hémisphères cérébraux ; et ce malgré que tous deux seraient qualitativement identiques.

Pour Snowdon et Olson, si seulement la moitié des personnes humaines (les « pures-personnes ») avait un critère d'identité psychologique, cela contredirait le présupposé lockéen que toute personne doit, en vertu de sa vie mentale, avoir un critère d'identité psychologique. De plus, la possibilité « d'animaux-personnes » ayant à la fois une vie mentale comme la nôtre et des conditions de persistance biologiques conteste la pertinence même de la distinction lockéenne entre personne et animal : « if [the human animal] is a person, then persons cannot be things which, in their natures, are distinct

from animals. » (Snowdon 1990 : 94) Ainsi, si être une personne n'est pas un qualificatif restreint aux seules entités qui seraient essentiellement des personnes durant toute leur existence, mais est également applicable aux animaux humains, alors les raisons de croire que les personnes sont un type fondamental d'entité qu'il faut distinguer des animaux humains pourrait bien être erronée – c'est d'ailleurs là le point central de la solution animaliste au problème, dont nous allons discuter par la suite (2.1.3).

Par contre, si c'est plutôt la définition lockéenne de personne qui est trop permissive et qui ne devrait pas s'appliquer aux animaux humains, afin de maintenir la distinction entre ces deux sortes d'entités, alors seule la moitié des individus intelligents et conscients d'eux-mêmes compterait comme des personnes (Olson 2002a : 190). Bref, peu importe quelle branche du dilemme nous empruntons, la distinction lockéenne serait incohérente : soit nous refusons sans bonne raison le statut de personne à la moitié des êtres intelligents ; soit, si nous l'accordons à tous, nous affirmons arbitrairement que seule la moitié des personnes a des conditions de persistance psychologiques. C'est ce qu'Olson nomme le *Personhood Problem*.

Ces multiples problèmes constituent ensemble l'argument de l'animal pensant, dont la force est non-négligeable. Quelles réponses est-il possible de fournir à un tel argument ?

On pourrait croire que le réductionnisme constitutif de Parfit serait en mesure de fournir une réponse à ces objections en raison de son réductionnisme ontologique de la personne. En effet, selon cette position, nous, les personnes, serions des entités seulement *conceptuellement* distinctes de notre corps, de notre cerveau, et des expériences que nous vivons plutôt que des entités matérielles ou spirituelles supplémentaires dans le réel. Plus précisément, notre existence se réduirait à l'existence de notre corps, de notre cerveau et de nos expériences, ces éléments constituant notre existence, de sorte que nous n'aurions pas d'existence propre. Ce ne serait que parce qu'on affirme usuellement être *ce qui* a un corps, un cerveau et qui vit des expériences que nous serions alors un *bundle* conceptuel constitué de ces éléments mais distinct d'eux. On pourrait alors s'attendre à ce que la personne ne figure pas comme penseur additionnel, puisqu'elle n'est pas conçue comme une entité réelle, évitant ainsi le problème de dédoublement des penseurs. Ce n'est cependant pas si simple, parce que cette position soutient tout de même que la personne

est numériquement distincte de son support matériel. Un tel constitutionnalisme ne peut alors être défendu contre l'assaut des animalistes, et Parfit se voit contraint de l'abandonner. Tâchons de reconstituer la réflexion qui le mène à cette conclusion.

Pour ce faire, il nous faut d'abord reformuler l'argument de l'animal pensant. Le problème qu'il met en évidence, c'est que le *principe de la carrière* justifiant la distinction numérique entre la personne et l'animal – puisque leurs carrières pourraient diverger – se confronte à un autre principe lockéen, le *principe d'exclusion* :

« For we never finding, nor conceiving it possible, that two things of the same kind should exist in the same place at the same time, we rightly conclude, that whatever exists anywhere at any time excludes all of the same kind, and is there it self alone. » (Locke 1975 : II.XXVII, §1)

La raison de cette exclusion est la suivante. Si deux ou plusieurs objets du même type occupaient le même espace et partageaient la même matière, un même objet en serait plusieurs, et nous aurions alors une contradiction conceptuelle, anéantissant les conditions de possibilité mêmes d'identifier des objets et de faire des distinctions :

« [T]he Notions and Names of Identity and Diversity would be in vain, and there could be no such distinctions of Substances, or any thing else one from another. » (Locke 1975 : II.XXVII, §2)

Par suite, seules des substances de types différents pourraient coexister dans le même espace. Locke en dénombre trois : Dieu, les êtres dont l'intelligence est finie, et les corps – ou plus précisément les particules et par extension les corps qu'ils constituent (Locke 1975 : II.XXVII, §2). Il serait donc seulement cohérent de distinguer numériquement deux entités constituées de la même matière si elles sont de type différent.

Or, le constitutionnalisme initial de Parfit, qui distingue numériquement la personne de son corps, de son cerveau, et de ses expériences, n'était pas problématique, puisque ce sont là des types d'objets différents, voire pas des objets du tout dans le cas des expériences. Parfit en fait lui-même la remarque : « since these are objects of different kinds, we can explain why they are distinct. They have different criteria of identity. » (Parfit 2008 : 195) Le *principe d'exclusion* ne s'applique pas dans un tel cas. Parfit pouvait donc légitimement défendre, contre les partisans d'un critère corporel de l'identité personnelle tels Bernard Williams et Judith Jarvis Thomson, que corps et personnes sont des entités numériquement distinctes pouvant avoir des carrières différentes. Le corps persiste en effet comme cadavre suite à la mort de la personne et de

l'animal.

Cependant, le constitutionnalisme de Parfit se heurte au *principe d'exclusion* lorsqu'on tente de l'appliquer à l'animal, parce que l'argument de l'animal pensant montre qu'un animal et une personne sont des entités du même type, c'est-à-dire des êtres rationnels et conscients de soi. Parfit reconnaît le problème :

« We are already inclined to deny that people are their bodies; and, to defend that distinction, we need only claim that two different objects could share the same matter. We are less inclined to distinguish between persons and human beings; and, to defend that distinction, we would have to claim that two different conscious beings could share all of their thoughts and experiences. » (Parfit 2008 : 200)

Son constitutionnalisme, même s'il est réductionniste, n'est donc pas une manière facile de s'extirper du problème de l'animal pensant et d'affirmer que nous sommes un objet matériel, en l'occurrence un animal, et que nous avons commencé notre existence en tant que fœtus.

On pourrait toutefois répliquer avec la question suivante : si certes la personne est conceptuellement distincte de l'animal, en quoi devrait-elle être également numériquement distincte de lui ? La personne n'est-elle pas supposée se réduire à ses éléments constitutifs, en l'occurrence à l'animal ? et éviter ainsi le problème de l'animal pensant ? Parfit n'est pas explicite au sujet de ce qui cloche dans sa position initiale une fois qu'on l'applique à l'animal, mais nous pouvons présumer ceci. Comme les autres partisans de l'approche psychologique, il adhère à la *Transplant Intuition*, et propose des expériences de pensées analogues pour défendre sa position. Il croit que la personne pourrait survivre à la mort de l'organisme, ce qui implique qu'elle est une entité numériquement distincte de ce dernier. Or, puisque Parfit souscrit à la définition lockéenne (et cartésienne) de la personne comme entité essentiellement pensante, il doit alors affirmer que, même lorsque la personne est séparée de l'animal pensant, sa nature demeure celle d'une entité fondamentalement pensante. Mark Johnston est de cet avis :

« Since a person is a thinker, the friend of the Wide Psychological View, who treats a person as a bundle of mental events and states, had better say that such a bundle can think. » (Johnston 2007 : 58)

La position de Parfit demeure donc sujette à l'argument de l'animal pensant.

On peut comprendre que s'il y a tout de même un dédoublement des penseurs

chez Parfit, c'est parce que la réduction de la personne n'est pas complète. La personne n'est pas seulement le corps ou l'animal sous un autre nom, tel que le sont Hespérus et Phosphorus. Son *existence* est certes réduite à l'existence d'éléments constitutifs, la personne ne pouvant exister indépendamment d'un support matériel, mais l'*entité* qu'elle est n'est pas réduite au corps ou à l'animal, et pourrait survivre à leur mort. Il semble ainsi y avoir une incohérence irrésolue dans la position initiale de Parfit. Elle se trouve dans son présupposé que la personne est à la fois une entité conceptuelle existant seulement dans notre façon de parler, autrement dit qu'elle n'est pas le type d'entité capable de poser des actions par elle-même ; *et* qu'elle est essentiellement capable de pensée autonome. Un tel réductionnisme manqué s'apparente davantage à un dualisme, comme le soutient McDowell (voir 1.4).

D'ailleurs, le constitutionnalisme réductionniste de Parfit était une manière pour lui de justifier que malgré que nous n'existons pas indépendamment des éléments qui nous constituent, nous en sommes tout de même numériquement distincts. Si Parfit est cohérent, il ne peut prétendre l'utiliser de nouveau, cette fois-ci pour affirmer que malgré que nous puissions survivre à la mort de l'animal, autrement dit malgré que nous en sommes distincts, nous *sommes* des animaux et avons été des fœtus. Le constitutionnalisme ne peut à la fois faire de nous des entités numériquement distinctes de l'animal *et* numériquement nous y identifier.

En somme, l'argument de l'animal pensant ouvre grand la porte au *principe d'exclusion*. Deux êtres pensants ne peuvent occuper le même espace et partager la même matière, ce qui force les néo-lockéens à revoir à neuf leur position, Parfit y compris. Ne pouvant être numériquement identiques qu'à une seule entité, nous sommes alors placés devant un dilemme. Personne ou animal ?

Plusieurs stratégies sont possibles pour répondre à l'argument de l'animal pensant et construire une position à laquelle le principe de l'exclusion ne s'applique pas. Une première option consiste à affirmer qu'entre personne et animal, il n'y a qu'une seule entité composée de notre matière, et donc un seul penseur. Cette entité pensante, à laquelle nous sommes numériquement identiques, serait alors l'animal si nous voulons pouvoir affirmer que nous avons commencé notre existence en tant que fœtus. Nous adopterions alors une approche pleinement réductionniste par rapport à la

personne. Celle-ci ne serait pas numériquement distincte de l'animal, puisqu'elle ne serait simplement pas une entité, ou encore une entité seulement conceptuelle, sans capacité de penser qui lui serait propre. La distinction lockéenne serait donc abandonnée. C'est la solution des animalistes, que nous allons voir en premier (2.1.3), et Parfit l'embrasse à contre-cœur pour formuler sa seconde position, quoiqu'il le fait en conservant son approche psychologique de l'identité personnelle (2.2).

Une deuxième option consiste à maintenir qu'il y a chez chaque être humain deux entités numériquement distinctes constituées de la même matière, l'animal et la personne, mais qu'elles sont de types différents. Une seule de ces deux entités serait un être pensant, la personne suivant la définition lockéenne, et nous y serions numériquement identiques. Il faudrait donc refuser pensée et conscience à l'animal. On éviterait ainsi le *principe d'exclusion* tout en préservant la distinction lockéenne. C'est la position que défend Shoemaker depuis sa rencontre avec l'argument de l'animal pensant, intégrant les conséquences de l'argument à son constitutionnalisme. Parfit se rangera finalement à cet avis, quoique sa troisième et dernière position prend une forme toute différente. Alors que Shoemaker défend que nous sommes des êtres purement psychologiques mais matériellement instanciés (2.3.1), Parfit propose que nous sommes la partie pensante de l'animal (2.3.3).

2.1.3 – La réponse animaliste : la personne comme *sortal de phase*

Voyons d'abord la réponse animaliste classique quant à la sorte d'entité que nous sommes. Le problème que mettent en évidence les arguments précédents, c'est que les lockéens supposent que la personne est un type d'entité fondamental plutôt qu'une manière de décrire des entités ayant un ensemble de propriétés mentales particulières, tel que c'est le cas des animaux humains durant une partie de leur vie. Nous allons d'abord voir comment les animalistes justifient leur position, et ensuite ce que ça implique quant à notre identité animale et personnelle.

Pour clarifier l'idée que la personne humaine ne serait que l'animal humain saisi sous une certaine description, Snowdon, et Olson à sa suite, s'appuient sur la théorie de l'individuation de David Wiggins. Selon Wiggins, pour tout objet identifiable, il y a un concept *sortal* qui répond à la question « *qu'est-ce que c'est ?* » et qui l'individue

(Wiggins 2001 : 64). Un tel sortal, valable durant toute la carrière d'un objet et sans lequel cet objet n'existerait plus, est ce qu'il nomme un *sortal de substance* (*substance concept*) :

« f is a substance-concept only if the grasp of f determines [...] what changes x tolerates without there ceasing to exist such a thing as x [...] » (Wiggins 2001 : 70)

En d'autres termes, un sortal de substance identifie un objet par les propriétés essentielles de la classe à laquelle il appartient et détermine quels changements accidentels il peut subir sans devenir numériquement différent. C'est d'ailleurs Locke qui nomme *sortaux* les concepts que nous utilisons pour isoler conceptuellement des entités afin de les classifier (Locke 1975 : III.III, §17), et la théorie de Wiggins le suit en cet aspect.

Selon Locke et Wiggins, le sortal de substance auquel appartient une entité détermine ses conditions de persistance et oriente alors nos jugements d'identité. En effet, pour porter un jugement d'identité, il faut préalablement être en mesure d'isoler (*single out*) des objets dans l'espace-temps. Cette opération de sélection d'éléments de notre expérience sensible pour les constituer en *un* objet distinct dépend à son tour de la sorte de chose en question, c'est-à-dire de notre manière de la classifier (Wiggins 2005 : 442). Par suite, puisque la délimitation d'un objet dans l'espace-temps dépend du sortal de substance qu'on lui applique, c'est ce même sortal qui détermine ce qui est requis pour que cet objet persiste. C'est une affirmation récurrente chez Locke, qui sert même de conclusion à son chapitre sur l'identité :

« [W]hatever be the composition whereof the complex *Idea* is made, whenever Existence makes it one particular thing under any denomination, the same Existence continued preserves it the same individual under the same denomination. » (Locke 1975 : II.XXVII, §29)

Jonathan Bennett modernise et clarifie ce passage ainsi :

« [T]he complex idea we use when classifying a thing as being of a certain kind also determines what it is for a thing of that kind to continue in existence. » (Locke 2004 : II.XXVII, §29)

Selon Locke, ce serait alors simplement parce que nos sortaux de substance ne sont pas clairement définis, et qu'on ne se tient pas à nos définitions lorsqu'elles sont claires, que nos jugements d'identité, à propos des personnes notamment, seraient incertains ou empreints de doute (Locke 1975 : II.XXVII, §28).

C'est donc parce que les lockéens traitent le concept de *personne* comme un *sortal* de substance, au sens où toute personne en serait une durant toute sa carrière, qu'ils croient que les personnes sont une sorte d'entité numériquement distincte d'autres sortes d'entités, ayant des conditions de persistance qui lui sont propres. Les expériences de pensée de Locke montrent certes que l'existence de la personne ne peut être rattachée à une *substance* en raison de la possible divergence de leurs carrières, mais cette indépendance au moins conceptuelle de la personne apparaît alors comme une raison de croire que la personne est une sorte d'entité à son propre titre, un *sortal de substance* distinct de l'animal :

« [A]ny two abstract *Ideas*, that in any part vary one from another, with two distinct names annexed to them, constitute two distinct sorts, or, if you please, *Species*, as essentially different as any two of the most remote or opposite in the World. » (Locke 1975 : III.III, §14)

De plus, c'est parce que nous individuons conceptuellement les personnes par leurs capacités mentales que le critère d'identité de toute personne doit être psychologique ; le *sortal* de substance déterminant leurs conditions de persistance. Olson remarque d'ailleurs que le seul fait de saisir la personne comme un *sortal* de substance conduit nécessairement à adopter une approche psychologique de l'identité personnelle (Olson 1997 : 29).

Pendant, comme le montrent les arguments animalistes présentés précédemment, croire que *personne* et *animal* sont tous deux des *sortaux* de substance est hautement problématique. Cette manière de concevoir les choses n'est toutefois pas la seule qui soit admissible. Snowdon souligne que la définition lockéenne de *personne* comme entité rationnelle capable de conscience de soi est seulement fonctionnelle : « Locke did not restrict in any way the sort of thing that can be a person, except as a thing possessing the specified set of mental capacities. » (Snowdon 1990 : 101) Cette définition n'implique pas que *personne* est un concept délimitant une sorte d'entité. Elle ne répond pas à la question « *qu'est-ce que c'est ?* », mais nous informe plutôt sur ce que l'entité peut faire, sur ses capacités (Olson 1997 : 32).¹²

Il semble alors tout à fait plausible que les capacités mentales retenues par Locke

¹² La distinction entre ce qu'une entité est et ce qu'elle peut faire a toutefois ses critiques. Les concepts d'*animal* et d'*organisme* seraient tout autant des concepts fonctionnels que celui de *personne* (voir Shoemaker 1999a ; Nichols 2010). Nous n'élaborerons pas davantage là-dessus ici.

puissent être acquises au cours du développement d'un individu, et puissent être subséquemment perdues, plutôt que de croire qu'il y a là deux individus dont les carrières divergent. Le concept de *personne* ne serait pas un sortal valable pendant toute la carrière d'une entité et ne se qualifierait donc pas comme sortal de substance. Il pourrait alors être un qualificatif applicable aux animaux humains durant la période où ils possèdent les capacités psychologiques en question. Snowdon conclut que la définition de *personne* donnée par Locke est insuffisante pour l'opposer à celle d'*animal*, toutes deux ne désignant pas nécessairement deux sortes d'entités mutuellement exclusives mais pouvant au contraire être compatibles.

La théorie de l'individuation de Wiggins aménage une place à des concepts sortaux valables seulement durant une partie de la carrière d'une entité, qu'il nomme des *sortaux de phase (phase sortals)* (Wiggins 2001 : 30). *Adolescent* et *philosophe* en sont des exemples, et, si les animalistes ont raison, *personne* le serait également. Selon eux, le sortal de substance qui s'applique à nous est *animal* ou encore *animal humain*, et c'est en acquérant les capacités mentales spécifiées par Locke que nous devenons des personnes durant une phase de notre vie. Un fœtus serait ainsi un être humain qui n'est pas encore une personne. Et si jamais nous perdions nos capacités de pensée rationnelle et de conscience de soi, cela signifierait seulement que notre existence en tant que personne se termine, pas qu'une entité meurt, contrairement à ce qu'implique la position néo-lockéenne :

« When someone lapses into a persistent vegetative state, his friends and relatives may conclude that his life no longer has any value. They may even conclude that he has ceased to exist *as a person*. But they don't ordinarily suppose that their loved one no longer exists at all, and that the living organism on the hospital bed is something numerically different from him. »
(Olson 2003a : 332)

En tant que sortal de phase, le concept de *personne* n'est pas non plus réservé à une seule espèce naturelle, mais serait applicable à une multitude d'entités – de sortaux de substance – ayant développé des capacités mentales supérieures.

Maintenant que nous avons une meilleure idée de la position animaliste au sujet de la sorte d'entité que nous sommes, voyons ce qu'elle implique quant à notre identité. Puisque les conditions de persistance d'une entité sont déterminées par la sorte d'entité fondamentale qu'elle est, nous, animaux, avons donc un critère d'identité biologique. Il y

a cependant une tension chez les animalistes quant au sortal de substance qui s'applique à nous, entre *animal* et *animal humain*. Suivant la théorie de Wiggins, Olson affirme que notre sortal de substance ne pourrait être l'*animal humain*, sans quoi nous n'aurions pas nos conditions de persistance *en vertu du fait* que nous sommes des animaux ; elles refléteraient plutôt ce qui fait de nous spécifiquement des *humains*, telles nos capacités à avoir une vie psychologique et à penser rationnellement (Olson 1997 : 111). Autrement dit, nous tiendrions alors nos conditions de persistance en vertu du fait que nous sommes des personnes ou des penseurs plutôt que des animaux, ce qui irait directement à l'encontre de la thèse animaliste. Olson soutient donc que c'est le concept d'*animal* compris comme *organisme vivant* qui joue le rôle de sortal de substance pour nous, ainsi que pour l'entièreté du règne animal. Tout animal aurait alors les mêmes conditions de persistance comme le défend Locke : la survie d'un animal requerrait la continuité de son organisme, c'est-à-dire la continuité de ses fonctions vitales essentielles (Olson 1997 : 135).

Wiggins, par contre, a récemment précisé qu'il penchait vers l'autre camp : « the persons we are have to be individuated [...] not under the generic concept *animal*, but under the specific sortal concept *human being*. » (Wiggins 2005 : 447) Il croit comme Locke que les meilleurs candidats au titre de sortaux de substance sont les noms d'espèces naturelles (Wiggins 2001 : 77). Mais contrairement à ce qu'en croit Olson, l'application du sortal spécifique d'*animal humain* à nous-mêmes n'a peut-être pas les implications qu'il redoute. Chez Locke en effet, alors même que le sortal de substance des espèces naturelles est le nom qui est attribué à chacune et qu'il détermine leur critère d'identité, tous les animaux (et les végétaux) ont le même critère biologique, soit la continuité de leur organisme, plutôt qu'un critère propre à leur espèce. Une théorie des concepts sortaux mériterait donc de clarifier en quoi les sortaux de substance déterminent les critères d'identité des individus. Une manière de réconcilier leurs positions serait d'avancer que l'espèce détermine ce qui est requis pour qu'une sorte d'organisme persiste, *précisant* le critère d'identité biologique général que partagent tous les êtres vivants.

Cela dit, puisque la *personne* est selon les animalistes un sortal de phase plutôt que de substance, elle n'aurait alors pas de critère d'identité propre :

« In a sense, then, there is no such thing as personal identity, any more than

there is such a thing as infant identity or philosopher identity. We can ask what it takes for an infant or a philosopher or a person to persist through time ; but it is a mistake to think that those beings have persistence conditions *as* infants or as philosophers—or, if I am right, as people. » (Olson 1997 : 27)

Le critère d'identité des personnes dépend du sortal de substance de l'entité ayant les capacités mentales en question, et ne serait pas unique pour toutes les personnes s'il est possible qu'il existe des personnes non-animales. Le critère d'identité d'une personne humaine, c'est donc son critère d'identité en tant qu'animal humain, et serait biologique. Il serait certes à la limite possible de parler d'un critère « d'identité » personnelle, quoique ça ne serait pas un critère d'identité numérique, propre à l'entité que nous sommes, mais plutôt le critère de notre identité *tant que ce sortal de phase s'applique*. Autrement dit, ce serait un critère d'identité relative, d'une entité sous une description particulière.

En ce sens, en réduisant la personne humaine à un sortal de phase de l'animal humain, l'animalisme poursuit et porte à son plein achèvement le réductionnisme de la personne et de l'identité personnelle dont Parfit s'est fait champion. Les personnes ne sont pas des entités, l'identité personnelle n'a pas d'importance intrinsèque, et même qu'à proprement parler il n'y a pas une telle chose que l'identité personnelle. Cela alors même que l'animalisme permet d'affirmer, conformément à nos intuitions, que *nous* sommes des entités matérielles, ayant une existence réelle plutôt que conceptuelle, contrairement à la position initiale de Parfit. Sous cette forme, la *ego theory* s'avère une version du réductionnisme plus robuste que la *bundle theory* de Parfit.

*

Portons un regard d'ensemble sur ce que nous avons vu dans cette section. L'argument du fœtus semble indiquer que l'approche psychologique de l'identité ne s'accorde pas avec l'affirmation que nous sommes des animaux. L'argument de l'animal pensant montre à son tour que la réponse néo-lockéenne à l'argument du fœtus, le constitutionnalisme, conduirait à un dédoublement des penseurs, de sorte que l'animal serait lui aussi une personne, ôtant ainsi toute pertinence à la distinction lockéenne entre personne et animal. Par conséquent, l'animalisme rejette cette distinction et propose que la personne ne serait qu'une manière de saisir l'animal. Il semble ainsi que seule une approche biologique nous permette d'être pleinement réductionniste à propos des

personnes et de maintenir que l'être humain n'est pas un mariage de deux entités. Cela devrait alors impliquer, si la position est cohérente, que nous ne sommes la même personne que tant que nous sommes le même animal, et, par suite, que la personne ne pourrait lui survivre.

2.2 – *Parfit II* : la récupération lockéenne des principes animalistes

Tournons-nous maintenant vers Parfit et sa réception des arguments animalistes. Il développe deux positions en réponse. Dans cette section, nous allons nous concentrer sur la première, c'est-à-dire sur sa deuxième position à propos de notre nature ontologique. Nous verrons l'autre, qu'il construit une vingtaine d'années plus tard, dans la section d'après (2.3).

Au premier contact avec l'animalisme, Parfit croit pouvoir conserver dans l'ensemble sa position initiale à propos de la personne et de l'identité personnelle, tout en affirmant désormais que nous sommes essentiellement des animaux, autrement dit que nous leur sommes numériquement identiques.¹³ Il affirme ceci : « The Animalist Argument, I conclude, does not undermine the view defended in my book. » (Parfit 2008 : 202) Ce n'est toutefois pas si simple, puisque bien des animalistes défendent comme Olson que notre vie mentale n'entre pas en compte dans notre identité (animale), alors que pour Parfit celle-ci est l'élément essentiel de l'identité personnelle. Il y a donc un sérieux travail à faire pour concilier les approches psychologique et biologique, et Parfit ne fait qu'en jeter les bases, sans jamais en défendre une version complète et cohérente.

D'ailleurs, seuls deux de ses textes traitent directement des arguments animalistes. Le premier, « *Persons, bodies, and human beings* », publié en 2008 mais écrit au début des années '90,¹⁴ quelques années après la parution de *Reasons and Persons*, nous sert de

13 Notons que cette manière d'énoncer la thèse animaliste principale a récemment été jugée problématique par Olson (voir son 2015b). Selon lui, dire que nous sommes *numériquement identiques* à des animaux est trop compliqué, et dire que nous sommes *essentiellement* ou *fondamentalement* des animaux est compatible avec un constitutionnalisme selon lequel nous sommes chacun constitués d'une entité qui est *essentiellement* ou *fondamentalement* un animal. Mais puisque nous n'avons pas de meilleure manière de contraster cette thèse avec celle que nous sommes avant tout des personnes, nous allons conserver ces formulations dans le mémoire.

14 Parfit ne semble pas certain de l'année où il a écrit ce texte. Dans le texte même, il parle de « around 1990 » (2008 : 207, note 27), alors que dans son 2012, il affirme l'avoir écrit « around

source principale ici puisque le deuxième, « *We are not human beings* », publié en 2012, expose plutôt sa position finale qui s'oppose nettement à l'animalisme. Nous allons tâcher de présenter la première réponse de Parfit à l'animalisme de la façon la plus charitable et cohérente possible, élaborant sur les pistes qu'il laisse tout en indiquant les ambivalences et incohérences qui demeurent irrésolues chez lui. Ces dernières expliquent d'ailleurs en partie pourquoi il abandonnera cette position mitoyenne pour revenir à un lockéanisme plus classique selon lequel nous ne sommes pas des animaux.

Pour ce faire, nous allons d'abord voir ce que Parfit emprunte à la position animaliste, notamment l'idée centrale que nous sommes des animaux (2.2.1). Il n'adopte toutefois pas l'animalisme parce qu'il lui trouve des problèmes majeurs, que nous allons ensuite examiner (2.2.2). Surtout, il veut conserver son approche psychologique de l'identité personnelle, et la seule position qui lui permettrait de le faire tout en affirmant que nous sommes des animaux consiste alors à considérer le cerveau comme l'organisme animal lui-même. Nous allons élaborer sur la maigre justification que Parfit propose en soutien à sa thèse (2.2.3). Finalement, nous allons tenter de dégager le critère de notre identité (animale) qui en découle, puisque Parfit ne le fait pas clairement lui-même (2.2.4).

2.2.1 – Le rapprochement animaliste de Parfit

Commençons par ce qui est le plus explicite dans la réaction de Parfit aux arguments animalistes, à savoir le fait qu'il prenne à l'animalisme l'idée que nous sommes numériquement identiques à des animaux. Nous allons nous attarder aux raisons qui motivent son choix, ainsi qu'à l'impact que cela a sur sa conception de la personne et de son identité.

D'abord, avant les premières publications de Snowdon et Olson en défense de l'animalisme, les positions de Parfit et de Shoemaker se recoupaient presque entièrement. Ce sont les arguments animalistes qui ont conduit les auteurs à se différencier, Parfit faisant un pas vers l'animalisme, alors que Shoemaker pousse plus avant son constitutionnalisme – quoique Parfit juge que ce ne sont là que des différences mineures. Le passage suivant le confirme :

1992 » (2012 : 8, note 7).

« Shoemaker's view differs slightly from mine. Shoemaker defends a pure version of the Psychological Criterion of personal identity, according to which some future person would be the same as some present person if and only if these persons would be uniquely psychologically continuous. Though I once defended this criterion, I wouldn't do so now. And Shoemaker assumes that what we are essentially is persons, while I regard it as acceptable to claim that what we are essentially is human beings, treating the concept *person* as a phased-sortal, like *child* or *chrysalis*, so that we exist before we become persons and we may continue to exist after we cease to be persons. » (Parfit 1999 : 218)

Nous allons traiter des points importants qui en ressortent tour à tour.

Le premier changement qu'on constate chez Parfit, c'est qu'il adopte le point principal des animalistes : nous sommes numériquement identiques à des animaux humains, et la personne n'est qu'un sortal de phase c'est-à-dire un descripteur qui s'applique à nous durant une période de notre existence et qui reflète le fait que nous possédons un ensemble de propriétés mentales particulières. L'argument du fœtus a fait mouche, et c'est donc avant tout afin de pouvoir affirmer que nous avons été des fœtus et pourrions perdre nos capacités mentales durant notre vie qu'il modifie sa position. Cela est en fort contraste avec sa position dans *Reasons and Persons* où il cherchait à convaincre ses lecteurs qu'ils avaient une fausse conception d'eux-mêmes ; que le réductionnisme à propos de la personne était « the truth about *ourselves* » (R&P : 280 ; c'est moi qui souligne) ; et que nous n'avions ainsi qu'une existence conceptuelle. Ce changement peut paraître mineur dans une discussion portant seulement sur la personne et le statut moral qui lui est associé, de sorte que si Parfit le mentionne dans ses articles ce n'est généralement qu'au passage. Cependant, se concevoir comme des *animaux* qui sont des personnes, ayant donc une existence réelle et matérielle, plutôt que comme des *personnes* qui ne sont pas des animaux, n'existant que comme artifices langagiers, représente une transformation radicale.

Pour pouvoir affirmer ainsi que nous sommes des animaux, Parfit doit alors laisser tomber la distinction lockéenne selon laquelle personne et animal sont numériquement distincts. Contrairement à ce qu'on pourrait croire, ce n'est pas la conséquence peu plausible de la distinction – à savoir que le corps de chaque être humain abriterait en fait deux personnes partageant toutes leurs pensées – qui le mène à la rejeter. La charge d'incohérence admet selon lui une réponse compatible avec le *principe*

d'exclusion :

« Even if human beings are also persons, they must be persons of a *different* kind from the ones who are not human beings. These two kinds of person have different criteria of identity. » (Parfit 2008 : 198)

Notre concept de *personne* serait en fait deux sortaux de substance distincts l'un de l'autre, l'*animal-personne* et la *pure-personne*, s'appliquant à deux entités de types différents. Le fait qu'elles soient toutes deux conscientes et rationnelles n'en ferait donc pas nécessairement des entités du même type ; elles pourraient ainsi se soustraire au *principe d'exclusion* lockéen.

Cependant, Parfit ne spécifie pas en quoi ces personnes diffèrent hormis par la possibilité qu'elles ne partagent pas la même matière durant toute leur carrière. Justifier ainsi l'idée qu'elles seraient des entités de *types* différents malgré qu'elles soient qualitativement identiques nous semble davantage une manière d'ignorer le *principe d'exclusion* que de le respecter. De plus, même si cette solution était acceptable, elle ne serait que partielle puisqu'il faudrait alors justifier comment un seul événement mental pourrait avoir deux sujets. Parfit admet qu'il n'a pas de réponse satisfaisante à cette question, mais considère que l'argument de l'animal pensant est tout de même insuffisant pour montrer que la distinction lockéenne est incohérente (Parfit 2008 : 198-9).

Ce que cette réponse de Parfit indique à notre avis, c'est qu'il faut cerner précisément ce qui nous permet d'affirmer qu'une entité est d'un type différent d'une autre, sans quoi l'application du *principe d'exclusion* sera arbitraire. Si l'on se fie à Locke ou encore à l'ontologie de van Inwagen que nous avons évoquée plus tôt (1.2.2), les particules élémentaires et les êtres intelligents (vivants pour van Inwagen) seraient les deux types d'entité qui pourraient coïncider et qui excluraient l'existence d'autres entités du même type dans le même espace – si l'on fait abstraction de Dieu. Il ne pourrait alors y avoir qu'un être capable de pensée dans un même espace, et il serait acceptable de diviser *conceptuellement* le corps humain en autant de parties superposables que nous désirons, tant que nous ne supposons pas que plus d'une particule du même type occupe le même espace.

Cela dit, le défaut que Parfit trouve à la distinction lockéenne, ce n'est pas qu'elle est incohérente, mais plutôt que l'argument de l'animal pensant montre qu'elle n'est d'aucune utilité : « The Lockean distinction cannot be useful, though, except in the cases

where it makes a difference. And, in these cases, it is no use. » (Parfit 2008 : 199) Dans la majeure partie des cas, il ne résulterait de la distinction qu'une ambiguïté bénigne de nos pensées : elles auraient la même valeur de vérité peu importe si elles étaient appliquées à l'animal-personne ou à la pure-personne (Parfit 2008 : 199). Dans les rares cas où l'attribution à l'une ou l'autre changerait leur valeur de vérité, par exemple lorsque j'affirme que « je suis un animal », la distinction lockéenne ne nous aiderait en aucune manière à déterminer le référent du « je », comme l'a indiqué Olson (2.1.2). Parfit la rejette donc.

Par contre, si distinguer *numériquement* les personnes des animaux humains ne saurait convenir, il reste malgré tout des raisons de les distinguer *conceptuellement* ; Parfit et les animalistes le reconnaissent. Les animaux humains ne sont pas des personnes durant toute leur existence d'une part ; et il pourrait y avoir des personnes de nature autre qu'humaine d'autre part, de sorte que ces concepts sont considérés comme des sortaux de types différents, de substance et de phase, comme nous l'avons vu. Cependant, conformément à sa position initiale et contrairement à la position animaliste classique, la personne demeure selon Parfit une *entité* distincte de l'animal même si elle n'en est plus numériquement distincte. Cela lui permet de continuer à traiter la personne indépendamment de l'animal et de parler d'un critère d'identité *personnelle*, plutôt que de dire comme Olson que le critère d'identité des personnes humaines n'est que celui des animaux humains. Une bonne partie de ce que Parfit aurait précédemment avancé à propos de la personne et de son identité ne serait donc pas invalidé. Cela comporte toutefois des problèmes puisque Parfit continue de traiter le critère d'identité personnelle comme critère d'identité numérique, comme si la personne ne se réduisait pas à l'animal ; alors que ce serait plutôt un critère d'identité relative, à propos de l'animal humain *sous une description particulière*. À notre avis, cela trahit une incohérence dans sa position. Il ne serait pas parvenu à se défaire pleinement de l'emprise de la distinction lockéenne – qu'il réintroduira d'ailleurs dans sa troisième position.

Ensuite, la deuxième chose à noter dans la citation où Parfit se compare à Shoemaker, c'est qu'il change de position quant au critère d'identité de la personne. Jusqu'à la première édition de *Reasons and Persons* inclusivement, Parfit défendait comme Shoemaker un critère purement psychologique d'identité personnelle (*The Wide*

Psychological Criterion). Pour être identique à une personne future, il fallait seulement que nous soyons en continuité psychologique avec elle et que cette relation ne soit pas dédoublée. Ce n'est qu'en 1987, lors de la seconde édition de *R&P*, qu'il se rétracte et affirme que nous n'avons pas besoin de choisir entre les multiples versions du critère d'identité psychologique (*R&P* : *x*). Les autres versions ajoutent une clause additionnelle : que la continuité psychologique soit produite par une cause fiable, ou encore par sa cause normale (la continuité de suffisamment du même cerveau), plutôt que par n'importe quelle cause. Ce qui justifie son indifférence quant à la version du critère, c'est que même si le même cerveau était nécessaire pour préserver notre identité personnelle, la continuation de notre vie mentale est ce qui importe principalement pour nous et devrait à son avis être presque aussi satisfaisante que notre survie (*R&P* : 209).

Cependant, malgré que Parfit continue d'affirmer que la version du critère importe peu dans maints de ses articles subséquents, il défend clairement le critère psychologique requérant le même cerveau (*The Narrow Psychological Criterion*). Avant de changer de position une troisième et dernière fois en 2012, il l'admet :

« The Lockean view that I have earlier defended, which I called the *Narrow, Brain-Based Psychological Criterion* states: If some future person would be uniquely psychologically continuous with me as I am now, and this continuity would have its normal cause, enough of the same brain, this person would be me. » (Parfit 2012 : 6)

Parfit n'explique pas les raisons pour lesquelles le cerveau gagne graduellement de l'importance dans sa conception de la personne, mais nous pouvons fournir l'interprétation suivante: sa volonté d'affirmer que nous sommes des animaux le conduit à rattacher le critère d'identité de la personne à celui de l'animal. Si nous n'avons pas tort, le cerveau est ce qui ancrerait alors la personne à l'animal. De plus, cela éviterait que sa position soit interprétée comme réifiant nos expériences et impliquant qu'elles pourraient exister par elles-mêmes (voir 1.4).

Bref, en élaborant sa deuxième position, Parfit rejette comme les animalistes la distinction lockéenne ; il concède que nous sommes numériquement identiques à des animaux et que les personnes ne sont que des sortaux de phase. Il n'adopte cependant pas en bloc la position animaliste classique. En effet, il rejette l'idée que le critère d'identité personnelle serait biologique, optant plutôt pour rajouter une clause biologique au critère

psychologique lockéen. S'il préfère cette approche, c'est non seulement parce que notre vie mentale est ce qui a la plus grande importance à nos yeux, mais aussi parce qu'il y a des problèmes majeurs dans la position animaliste – que nous allons voir à l'instant.

2.2.2 – Les problèmes de l'animalisme classique

Examinons maintenant les problèmes que Parfit voit dans la position animaliste ; ceux-ci lui servent de raisons pour l'écartier et construire la sienne bien qu'il conserve l'idée que nous sommes des animaux. Selon Parfit, l'animalisme classique balaie trop vite la *Transplant Intuition* – la croyance que nous irions là où iraient nos hémisphères cérébraux parce qu'ils seraient le support de notre vie mentale – sans en considérer toutes les conséquences. Une difficulté majeure pour l'animalisme est ce que Mark Johnston nomme le *problème des « personnes-résiduelles »* (*The Remnant-Person Problem*). Olson le voyait venir en développant ses arguments (Olson 1997 : 120-121), mais c'est Johnston (2007) qui lui donne le plein développement, que Parfit reprend. Ce problème montre que l'animalisme classique échoue à réduire pleinement la personne à l'animal, ce qui implique alors que l'animalisme serait sujet aux mêmes conséquences ontologiques inacceptables qu'il reproche au constitutionnalisme, notamment le dédoublement des penseurs.

Revenons sur l'expérience de pensée de la transplantation des hémisphères. L'animalisme classique défend que, contrairement à la *Transplant Intuition*, nous n'irions pas là où iraient nos hémisphères, parce que nous sommes numériquement identiques à un organisme plutôt qu'à une partie d'organe ou à la vie mentale qu'il supporte. Un organisme animal peut survivre à une transplantation d'organe, et la vie mentale de cet animal n'est pas nécessaire à son existence selon cette position.

Imaginons maintenant que contrairement à l'expérience de pensée originale où nos hémisphères étaient immédiatement transplantés dans un autre organisme (voir 2.1.1), ceux-ci seraient plutôt placés dans une cuve capable de les maintenir artificiellement en fonction. Si l'on suppose que les hémisphères suffisent pour supporter la conscience, nos hémisphères demeureraient capables de pensée rationnelle et de conscience de soi. Johnston remarque qu'ils se qualifieraient alors comme une personne selon la définition lockéenne (Johnston 2007 : 46). Le problème pour les animalistes,

c'est que si la personne pouvait effectivement survivre dans une cuve séparément d'un organisme animal, elle ne pourrait être conçue comme un ensemble de propriétés appartenant à cet animal. Séparer les hémisphères de l'organisme humain créerait donc une nouvelle entité, une *personne-résiduelle*, numériquement distincte de l'animal original. Si on imagine le problème autrement, le seul fait de détruire les organes vitaux non-nécessaires pour supporter notre vie mentale, même si on y substituait des équivalents fonctionnels, provoquerait notre mort et la création d'un *nouvel* individu non-animal, malgré qu'il conserverait notre vie mentale. L'animalisme classique ne réduirait donc pas pleinement la personne humaine à un sortal de phase de l'animal.

Ainsi, l'argument de Snowdon selon lequel un animal pensant se qualifie comme une personne selon la définition lockéenne, et qu'alors la personne ne pourrait être un sortal de substance mais serait plutôt un sortal de phase, est retourné contre l'animalisme. Si un organe ou une partie de l'animal se qualifie comme une personne mais pas comme un animal, cela suppose que la personne ne pourrait être un sortal de phase et serait plutôt un sortal de substance, un type d'entité fondamental. Affirmer que la continuité psychologique importe peu pour l'identité des animaux, comme le faisait Olson pour contrer la *Transplant Intuition*, s'avère insuffisant pour contrer l'objection des personnes-résiduelles.

Johnston avance que les animalistes n'ont alors que deux options pour rendre compte des personnes-résiduelles (Johnston 2007 : 47) : soit ils défendent que leur existence est impossible, soit ils doivent admettre que la séparation des hémisphères d'avec l'organisme crée une nouvelle entité, une pure-personne – ce qui revient à adopter la distinction lockéenne. C'est la raison principale pour laquelle Johnston rejette l'animalisme malgré qu'il croit, comme le second Parfit, qu'il est absurde de nier que nous sommes des animaux. Il y a certes une troisième option pour résoudre le problème : affirmer que les hémisphères cérébraux *sont* le même animal qu'avant l'opération, quoique cela requerrait de justifier l'idée qu'ils seraient un organisme à part entière plutôt qu'une partie d'organe. Mais nous anticipons.

Parfit élabore sur l'objection de Johnston afin de montrer que si la distinction lockéenne refait ainsi surface, le problème du dédoublement des penseurs ne saurait être loin derrière. Supposons que l'opération ait été effectuée avec succès et que nous ayons

deux hémisphères fonctionnels dans une cuve, autrement dit une personne non-animale. Nous voulons redonner à cette pure-personne un corps, de sorte que nous nous arrangeons pour entrer en possession d'un organisme humain, vivant et apprêté pour recevoir les hémisphères. Nous avons donc là deux êtres vivants, qui, suite à la transplantation, ne seront hypothétiquement qu'un seul individu, partageant la même vie. Que se passera-t-il alors avec la personne non-animale, une fois que l'animal aura reçu ses hémisphères et gagné les capacités mentales qui font de lui une personne ?

La situation est l'inverse de celle du fœtus ou du jeune bébé devenant une personne durant son développement, que nous avons vue plus tôt (2.1.1). Il aurait été bien pratique pour la position lockéenne classique que le jeune animal disparaisse une fois ses capacités mentales développées et qu'il cède le contrôle de son corps à la personne ; mais c'était une proposition absurde, quoique son rejet contraignait les lockéens à admettre l'existence de deux entités distinctes mais coïncidentes. De manière analogue, il serait bien commode pour les animalistes que la personne-résiduelle disparaisse une fois ses hémisphères joints à l'animal et qu'elle lui laisse le contrôle de sa vie mentale, mais cela est tout autant absurde. Par contre, si la pure-personne ne disparaissait pas mais survivait à l'opération, *et* que l'organisme acquérait les capacités mentales qui font de lui une personne grâce à ses nouveaux hémisphères, il y aurait alors chez le même être humain deux personnes partageant toutes leurs pensées (Parfit 2008 : 201). Par conséquent, l'animalisme se trouve confronté à sa propre version des problèmes de la multiplication des penseurs, des personnes, des vies, ainsi que des entités. Le *principe d'exclusion* n'est pas davantage respecté par l'animalisme que par le lockéanisme. L'animalisme classique, qui se présentait comme solution aux problèmes identifiés dans le lockéanisme tout en étant compatible avec la science, perd ainsi une partie de son attrait.

En somme, l'animalisme classique, en plus de contredire la *Transplant Intuition* et de ne savoir rendre compte des personnes-résiduelles, est sujet au même problème qu'il avait identifié dans l'animalisme, à savoir la multiplication des penseurs et personnes dans notre ontologie. Parfit ne l'adopte donc pas, mais propose plutôt une différente position qui lui permettrait d'avancer à la fois que nous sommes des animaux et que le critère d'identité personnelle est psychologique.

2.2.3 – Le cerveau comme organisme animal

Dans cette section, nous allons chercher à répondre à la question suivante : comment Parfit parvient-il à réconcilier l'idée que nous sommes numériquement identiques à un animal avec celle que les personnes humaines ont des conditions de persistance psychologiques ? Comme on va le voir, cela implique de considérer que le cerveau, porteur de notre vie mentale, *serait* l'organisme animal lorsque détaché du corps, plutôt que seulement un de ses organes. Pour qu'une telle position soit défendable, il faudrait que le cerveau joue un rôle vital, indispensable à la survie de l'organisme, ce que nous allons tâcher d'évaluer. Nous allons également voir une objection à laquelle s'expose cette position : le *problème de l'animal-résiduel* (*The Remnant-Animal Problem*), qui n'est que la contrepartie du *problème de la personne-résiduelle* qui afflige la position animaliste. Nous examinerons enfin comment l'ontologie de van Inwagen pourrait nous aider à le résoudre.

Prenons la question de départ comme un puzzle à résoudre afin de comprendre le raisonnement qui conduit Parfit à sa solution. Si nous voulons que, conformément à la *Transplant Intuition*, nous allions là où vont nos hémisphères cérébraux parce qu'ils assurent la continuité de notre vie mentale ; si nous voulons également éviter l'implication qu'une personne-résiduelle serait créée si nos hémisphères étaient séparés de notre organisme ; et ce tout en affirmant que nous sommes numériquement identiques à des animaux ; il semble que la seule avenue possible consiste à faire converger les critères d'identité de l'animal et de la personne, autrement dit à défendre l'idée que l'animal *et* la personne iraient là où vont nos hémisphères cérébraux. D'ailleurs, si la personne est bien un sortal de phase de l'animal, cela implique qu'elle ne survive que si l'animal survit, et qu'elle ne survive qu'en tant que cet animal. Il faudrait donc adopter une position selon laquelle l'animal *est* ses hémisphères cérébraux, ou du moins qu'il survit si et seulement si ceux-ci sont préservés. C'est la stratégie qu'emploie Parfit :

« [I]f the identity of animals partly depends on the importance of their different organs, our criterion must be sensitive to the kind of animal that we are considering. Not all animals have the same organs. [...] [I]n the case of animals like us, the brain is not merely one of several vital organs. Though it is only a single organ, it *is* special. It is the controlling organ. » (Parfit 2008 : 201) « [...] On that view, if our brains will continue to exist and function, we shall continue to exist. » (Parfit 2008 : 202)

Les animaux humains auraient donc un critère d'identité impliquant la continuité de suffisamment de leurs hémisphères cérébraux, à la fois parce que ces derniers contrôlent le reste du corps et parce qu'ils causent et assurent la continuité psychologique qui constitue notre identité personnelle. Le critère d'identité de l'animal humain se recoupe ainsi avec le critère d'identité de la personne humaine en ce qu'ils requièrent tous deux la continuité de nos hémisphères.

Là s'arrête le traitement que Parfit fait de sa seconde position, son intérêt se trouvant principalement dans les conséquences éthiques d'un réductionnisme de la personne. D'un point de vue ontologique, les implications et les interprétations d'une telle position sont multiples ; nous allons les traiter une à une.

Notons d'abord que cette seconde position de Parfit s'avère très similaire à la position de Thomas Nagel qu'il critiquait pourtant précédemment. Nagel cherchait pareillement à rattacher la personne à l'animal que nous sommes, défendant ceci :

« If [...] my mental life depends entirely on certain states and activities of my brain, and if some form of dual aspect theory is correct, then that brain in those states (not just in its physical states) is what I am, and my survival of the destruction of my brain is not conceivable. » (Nagel 1986 : 40)

Chez Nagel, la personne et l'animal que nous sommes étaient implicitement fondus en une seule entité : l'être humain avait une dimension objective, son corps physique, et une dimension subjective, la personne psychologique – toutes deux étant préservées par le cerveau.

Une des conséquences majeures d'une position qui soutient que l'animal humain serait essentiellement son cerveau, sur laquelle nous allons nous attarder ici, c'est qu'il lui faut revoir ce qu'on veut dire par *organisme* afin de pouvoir considérer le cerveau (ou les hémisphères) non pas comme un simple organe, mais comme un organisme à plein titre lorsque détaché du corps. L'alternative, soit de rejeter l'idée que l'animal humain serait un organisme, est hautement problématique puisqu'elle remettrait en question toutes nos connaissances en biologie. Si l'on juge cependant que le cerveau ne pourrait se qualifier comme un organisme, il serait alors plus plausible d'affirmer que nous ne sommes pas des animaux humains mais une *partie* de ceux-ci, plutôt que de dire qu'un animal n'est qu'un organe. C'est là d'ailleurs la troisième et dernière position de Parfit, que nous allons voir après celle-ci (2.3).

Qu'est-ce alors qu'un *organisme* ? Rappelons que Locke le définit comme une organisation fonctionnelle de parties capable de maintenir la vie. Les animalistes comme Olson comprennent cette définition comme voulant dire que la survie de l'organisme animal requiert la subsistance de ses organes vitaux essentiels. Ils admettent certes que certains de ces organes soient remplaçables puisqu'ils le sont en pratique, même le cœur, mais maintiennent qu'il y aurait tout de même *un* organe vital essentiel chez les êtres humains. Le cerveau serait-il cet organe ?

À ce sujet – et c'est une précision que nous avons omise jusqu'à présent – Olson et Parfit s'entendent pour dire que l'animal *et* la personne survivraient si le cerveau *entier* était maintenu en fonction séparément du reste du corps. La différence principale entre eux quant au cerveau, c'est que Parfit insiste sur l'importance des hémisphères parce qu'ils seraient la cause de notre vie mentale et la partie contrôlant notre corps, alors qu'Olson juge que seul le tronc cérébral est essentiel à notre survie parce que ce serait l'organe qui régule les fonctions vitales de l'organisme. C'est pourquoi les expériences de pensée qu'ils mobilisent et que nous avons présentées séparent ces deux parties du cerveau.

Par contre, Olson a récemment minimisé l'importance qu'il accordait au tronc cérébral, admettant que c'était une erreur de croire que la vie de l'organisme animal était rattachée à un seul « organe de maintenance » (Olson 2016). Il se base sur les données empiriques rapportées par le neurologue Alan Shewmon, qui montrent que l'unité de la vie de l'organisme n'est pas assurée par le cerveau, mais résulte d'un mécanisme holistique et non-localisable dans un seul organe :

« The integrative functions of the brain, important as they are for health and mental activity, are not strictly necessary for, much less constitute, the life of the organism as a whole. Somatic integration is not localized to any single “critical” organ but is a holistic phenomenon involving mutual interaction of all the parts. » (Shewmon 2001 : 473)

Si cette conclusion est correcte, la vie de l'organisme pourrait être maintenue sans aucune partie du cerveau, tronc cérébral compris. La mort cérébrale ne serait pas la mort de l'animal humain. Cela ne fait cependant que renforcer pour Olson l'insignifiance de nos hémisphères et de notre vie mentale pour notre survie.

Néanmoins, les résultats de Shewmon n'impliquent pas que la vie de l'organisme

ne pourrait être maintenue par le cerveau ou les hémisphères seuls ; au contraire, ils appuient tout autant la position de Parfit, puisque aucun organe vital ne serait essentiel pour préserver la vie. Shewmon, pour qui un organisme se définit comme une entité possédant au moins une propriété holistique, applicable au tout, est de l'avis qu'un cerveau conscient mais séparé de son corps se qualifie ainsi :

« [A]n isolated living brain, supporting retained consciousness, would have to be classified as possessing integrative unity (i.e., as being a live “organism as a whole,” although hardly a “whole organism” and in fact a severely mutilated and moribund one), precisely because the consciousness would represent at least one emergent, holistic-level property. » (Shewmon 2001 : 461)

Le différent entre Olson et Parfit demeure donc entier. Cependant, les expériences de pensée de transplantation, de mise en cuve et de décapitation visant à séparer la personne de l'animal seraient tout autant valides avec tout le cerveau qu'avec les seuls hémisphères. Bref, les résultats de Shewmon peuvent donc servir à justifier que le cerveau (ou les hémisphères) satisfasse la définition lockéenne d'*organisme* même lorsque séparé de son corps, parce qu'il pourrait préserver la vie.

Par contre, si Shewmon a raison et qu'à la fois le corps et le cerveau pourraient être des organismes, cela introduit un nouveau problème pour l'approche psychologique : il lui faudrait expliquer pourquoi nous survivrions en tant que l'organisme cérébral plutôt que l'organisme corporel, pour ainsi dire. Peu importe où on fait la coupe, si le cerveau ou les hémisphères demeuraient fonctionnels tout comme le reste du corps, moyennant un système artificiel de soutien de la vie pour chacun, il semble qu'il y aurait alors deux organismes qui survivent desquels tout au plus un seul pourrait être numériquement identique à l'animal que nous étions avant l'opération. Alors que l'animalisme classique générait le *problème de la personne-résiduelle* lorsqu'on séparait le cerveau de son corps, le second lockéanisme de Parfit a un problème analogue : celui de l'*animal-résiduel* (Olson 2015c), ce corps décérébré qui se qualifie également comme un organisme en plus du cerveau. Qu'est-ce qui nous permettrait de dire que nous, l'animal, serions notre cerveau ou nos hémisphères, plutôt que l'animal-résiduel ?

Pour répondre à une objection de ce type, Parfit nous renvoie à van Inwagen. Notons que ni Parfit ni van Inwagen ne prennent en compte les résultats de Shewmon puisque leurs écrits sont antérieurs. Ils anticipent tout de même la possibilité d'une double

survie biologique, et van Inwagen fournit des arguments pour la réfuter. Il défend qu'un tel animal-résiduel ne serait tout simplement pas un organisme animal, même s'il maintient certaines fonctions vitales ; s'opposant donc implicitement aux définitions de Locke et de Shewmon. Examinons sa position.

Rappelons que pour van Inwagen, comme nous l'avons vu dans le premier chapitre, il n'y a, outre les particules élémentaires, qu'un seul type d'entité naturellement individué ayant une existence réelle : les êtres vivants. Contrairement aux autres amas de particules, les vivants résultent d'un processus de *composition* de particules qui donne au tout des propriétés causales propres – holistiques pourrait-on dire, pour relier sa position à l'analyse de Shewmon. Selon van Inwagen, il n'y a pas de cerveau ni de corps ; ce ne sont là que des objets virtuels, conceptuellement individué. Lorsqu'on sépare un cerveau de son corps, nous avons en fait un organisme animal, constitué du cerveau, et un « complément-cérébral » (van Inwagen 1990 : 173). Ce complément aurait beau être maintenu « en fonction » grâce à un système de soutien de la vie, son activité ne constituerait pas une *vie* ; il ne serait rien de plus qu'un agrégat de « tissus vivants » (van Inwagen 1990 : 177).

Un de ses arguments, qui est d'ailleurs l'argument classique en faveur de la conception cérébro-centrique de l'organisme, est réfuté par Shewmon. Van Inwagen avance que le cerveau aurait pour fonction d'intégrer l'activité des cellules et des organes du corps en un tout coopératif (van Inwagen 1990 : 180), ou, comme le formule Shewmon, d'être « somatiquement intégrateur ». Les données semblent plutôt montrer que le cerveau n'est pas essentiel à une telle intégration, mais sert plutôt à la compléter et l'augmenter :

« [T]he brain's role here is one of modulating, fine-tuning, and enhancing an already well-functioning immune system, not of imperiously micromanaging a passive and basically incompetent immune system. The same could be said for all other somatically integrative functions: they are all the more effective when modulated by the brain, but they do not entirely vanish without the brain. » (Shewmon 2001 : 471)

D'ailleurs, certaines fonctions essentielles à la survie, la respiration et la circulation, ne sont pas médiées par le cerveau (Shewmon 2001 : 469). Shewmon conclut que quoique le cerveau *contribue* à l'unité vitale de l'organisme humain, il n'en est pas la source (Shewmon 2001 : 472).

Cependant, si l'intégration des fonctions vitales n'est pas le rôle du cerveau seul, il reste que ce n'est pas la seule fonction qui pourrait en faire la partie essentielle de l'organisme. La conception de l'*organisme animal* comme ensemble de fonctions vitales *somatiques* est peut-être réductrice étant donné que l'activité de celui-ci ne se résume pas au bon fonctionnement de ses organes. En effet, sans une activité dirigée vers le monde extérieur, un animal sans système de soutien de la vie trouverait la mort. Nous ne pouvons nullement épuiser la question ici, mais nous pouvons esquisser une réponse suivant la piste tracée par van Inwagen. Si nous comparons le cerveau au corps décérébré, leurs besoins pour qu'ils soient pleinement fonctionnels plutôt que simplement maintenus en « vie » sont grandement différents. Van Inwagen remarque qu'un cerveau installé dans un corps artificiel aura la possibilité de se maintenir lui-même, alors qu'un corps décérébré aura toujours besoin d'une intelligence externe qui le contrôle :

« Give the severed head the proper environment and it will maintain itself, but the headless body will need a constant supply of "instructions" in the form of electrically transmitted information. Unlike the head, it will not be able to coordinate its activities. A life-support system for the head will be no more than an elaborate pump. A mechanical life-support system for the headless body must involve the functional equivalent of a computer. » (Van Inwagen 1990 : 178)

Autrement dit, un organisme est une « unité autonome », un système qui fournit lui-même son principe d'organisation plutôt que de le tirer de l'extérieur, ainsi que le défendent les biologistes Humberto Maturana et Francisco Varela (1987 : 46-8).

Une telle manière de concevoir l'organisme va au-delà de l'idée d'un système continuellement entretenu par des mécanismes d'auto-régulation interne. Dans le cas des organismes animaux, la simple maintenance de leur existence requiert notamment qu'ils aillent chercher dans le monde extérieur les aliments nécessaires à leur subsistance. Pour être plus précis, ils sont motiles durant au moins une partie de leur existence ; les coraux par exemple perdent leur motilité lorsqu'ils se fixent aux sols marins. En ce sens, les animaux sont des entités dont la constitution nécessite que leur comportement soit orienté vers des objectifs se trouvant dans leur environnement extérieur. Autrement dit, une telle directionnalité est partie intégrante de l'organisme animal ; elle a un rôle *vital* central malgré qu'elle ne soit pas prise en compte dans la conception classique de l'organisme comme ensemble de fonctions vitales. Selon la conception proposée, les animaux seraient

essentiellement des êtres intentionnels (ou pré-intentionnels, selon la définition de l'intentionnalité adoptée). Il semble tout à fait plausible que le socle d'une telle faculté soit localisé dans le cerveau dans le cas des êtres humains et d'autres espèces similairement constituées.

Par suite, l'animal humain serait son cerveau plutôt que son corps décérébré, parce que cet organe serait non seulement capable de préserver la vie moyennant un système de soutien de la vie, mais également les facultés d'auto-organisation et de directionnalité (pré-)intentionnelle essentielles à l'organisme, sans parler de notre vie psychologique. Van Inwagen formule une idée qui va dans cette direction :

« [T]he brain is not only the seat of our experience and our conscious thought; it is also the seat of our capacity to have lives; it is the seat not only of one's intellect but of one's entelechy. » (van Inwagen 1990 : 179)

C'est là certes loin d'être un plaidoyer solide pour cette position ; notre objectif ici est simplement de donner une idée de ce que devrait défendre quelqu'un qui soutient, comme le second Parfit, que le cerveau ou les hémisphères *est* (la partie essentielle de) l'organisme humain.

En gros, afin de réconcilier l'idée que nous sommes des animaux avec l'approche psychologique de l'identité personnelle, tout en évitant les *problèmes de l'animal- et de la personne-résiduelle*, il faudrait considérer que le cerveau et non pas le corps décérébré serait l'organisme, tel que le suggère Parfit. Une telle position ne serait cohérente que s'il était possible d'attribuer au cerveau des fonctions vitales allant au-delà de celles que peut également maintenir le corps, autrement dit au-delà de la simple auto-régulation. En dire davantage requerrait des recherches empiriques. Nous suggérerons de tels appuis dans le troisième chapitre lorsque nous défendrons une position similaire à celle de Parfit mais animaliste (3.2).

2.2.4 – Les critères possibles de notre identité (animale)

Supposons maintenant que nous ayons une justification adéquate pour affirmer que l'organisme humain est essentiellement son cerveau (ou ses hémisphères). Quel serait alors le critère d'identité des animaux humains ? Nous savons que, pour Parfit, le critère d'identité *personnelle* est le critère psychologique *restreint*, qui requiert la continuité psychologique et suffisamment du même cerveau. Par contre, ce qu'il dit à propos de

nous, les animaux humains, est ouvert à au moins deux interprétations, quoique Parfit n'en adopte explicitement aucune. Examinons-les.

Une première possibilité serait d'avancer que nos critères d'identité en tant qu'animal et en tant que personne sont identiques. Autrement dit, le critère psychologique restreint vaudrait également pour l'animal humain. On pourrait croire que ce serait la position du second Parfit puisque la seule raison d'accorder de l'importance au cerveau vient selon lui de son rôle comme cause de notre vie mentale (R&P : 284). Le troisième Parfit affirmera d'ailleurs que croire que les hémisphères cérébraux pris isolément seraient l'organisme revient à adopter un critère psychologique de l'identité de l'animal humain. Il attribuera cependant cette position qu'il nomme *l'animalisme lockéen* à certains de ses opposants animalistes :

« If these Animalists now claimed that, in *Surviving Cerebrum*, the conscious rational being would be a living animal, who would be me, these people would be claiming that the true criterion of identity for developed human animals is of this Lockean psychological kind. » (Parfit 2012 : 12)

On pourrait voir dans l'attribution subséquente de cette position à ses opposants une répudiation de sa deuxième position, si c'est bien la sienne.

Ce ne serait d'ailleurs pas étonnant, puisque l'animalisme lockéen prête le flanc à de nombreuses objections. Il serait de nouveau sujet aux objections des animalistes, notamment à l'argument du fœtus. Les fœtus n'ayant pas de vie mentale ni de cerveau au tout début de leur existence, un tel critère impliquerait qu'ils ne seraient pas des organismes *humains* à proprement parler, quoiqu'ils seraient des organismes animaux. Autrement dit, si nous étions numériquement identiques à des animaux *humains* dont le critère d'identité est psychologique, *nous* n'aurions jamais été des fœtus. Et comme ce critère requiert la continuité de la série de nos expériences, nous ne serions pas le même *animal* si celle-ci était interrompue ou perturbée, si nous perdions la mémoire ou tombions dans un état végétatif par exemple.

Une telle position nous ramènerait ainsi à la case départ. L'élaboration conceptuelle rattachant l'être humain à son cerveau serait vaine puisque nous ne serions pas parvenus par ce moyen à identifier le concept sortal valable durant toute notre existence, mais n'aurions effectivement qu'ajouté un deuxième sortal de phase pour accompagner le concept de *personne*. Parfit considère cette possibilité de distinguer le

concept d'*être humain* de celui d'*animal humain*, le premier pouvant être conçu comme un sortal de phase dépendant de l'existence du cerveau, et le second comme le sortal de substance valable durant toute notre existence (Parfit 2008 : 204). Cette distinction ne serait cependant utile qu'afin de pouvoir dire que c'est en tant qu'êtres humains que nous avons un statut moral privilégié, plutôt que pour cerner la sorte d'entité que nous sommes.

Il y a certes une manière dont l'animalisme lockéen pourrait répondre à l'objection et considérer les fœtus comme des animaux humains malgré qu'ils ne possèdent pas encore toutes leurs facultés mentales. Il faudrait cependant réviser le critère psychologique pour également y intégrer la *capacité* à développer un cerveau et une vie mentale. *Nous* pourrions alors avoir été des fœtus. Par contre, une telle position irait tout de même contre l'esprit de la critique animaliste :

« When Animalists entered this debate, their main claim was that such psychological criteria of identity are seriously mistaken, because we are human animals, so that our criterion of identity must be biological. » (Parfit 2012 : 12)

Une position qui prend en compte les capacités de développement de l'organisme serait mieux servie par un critère biologique de l'identité de l'animal humain. Olson est également de cet avis : « If we were animals, we should have the identity conditions of animals. Those conditions would have nothing to do with psychological facts. » (Olson 2003a : 11) Certes, si effectivement des capacités (pré-)intentionnelles étaient vitales à l'organisme animal comme nous l'avons suggéré, la psychologie pourrait alors être considérée comme une extension de la biologie et y être intégrée. Ce que nous proposerons au troisième chapitre va dans ce sens (3.2).

Cela dit, si notre critère d'identité animale était biologique, on pourrait également le combiner avec un critère psychologique de l'identité personnelle – si l'on veut traiter comme Parfit la personne comme une *entité* conceptuellement distincte de l'animal plutôt que comme un prédicat. Attardons-nous à cette deuxième possibilité. Sous certains aspects, elle semble davantage compatible avec ce qu'affirme le second Parfit. En effet, si la personne est bien un sortal de phase applicable durant une partie de l'existence de l'animal humain, elle ne pourrait être la même que si elle est également le même animal humain. Il semble alors que le critère d'identité de la personne devrait être un sous-critère du critère d'identité de l'animal dont elle dépend. Plus précisément, le critère d'identité

personnelle serait le critère d'identité de l'animal humain, auquel on ajouterait une ou plusieurs clauses telle la continuité psychologique. Puisque le critère d'identité animale est biologique, le critère d'identité de la personne devrait donc être biologique *et* psychologique. Le critère d'identité personnelle restreint que Parfit défend, requérant la continuité de suffisamment du même cerveau et la continuité psychologique, correspond à ces exigences. Si nous y allons à rebours, le critère d'identité de l'animal humain devrait alors nécessiter la continuité de suffisamment du même cerveau.

Cette position à double critère semble prometteuse, et nous permettrait d'affirmer que nous, animaux humains, demeurerions numériquement les mêmes si notre vie mentale était affectée. Nous pourrions également affirmer que nous avons été un fœtus sans vie mentale si nous intégrons au critère biologique la capacité de développer un cerveau, comme nous l'avons précédemment mentionné. Notre critère d'identité animale serait donc ainsi formulé : la continuité de l'organisme humain serait la continuité de suffisamment de son cerveau *ou* de sa capacité à en développer un. Nous élaborerons là-dessus dans le troisième chapitre (3.2).

Cependant, Parfit n'assume pas les conséquences de cette position. Elle implique que si notre cerveau était détruit, nous, en tant qu'animal *et* personne, serions morts. Nous ne pourrions non plus survivre si l'on numérisait notre esprit et le transférons dans un clone de notre cerveau. Pour Parfit, par contre, notre identité numérique dans de tels cas est indéterminée ; il n'y aurait pas de réponse à la question « est-ce que je survis ? », quoiqu'on puisse arbitrairement en donner une (Parfit 2008 : 182-4). La position du second Parfit ne semble donc pas coller davantage à cette interprétation du critère d'identité de l'animal humain.

En somme, nous avons examiné deux critères d'identité de l'animal humain qui seraient compatibles avec la position du second Parfit selon laquelle la personne et l'animal iraient là où vont nos hémisphères cérébraux. Cependant, la première possibilité, l'animalisme lockéen, ramène les problèmes que l'animalisme reprochait au lockéanisme, alors que la position à double critère a des implications ontologiques que Parfit refuse. Ce résultat est peut-être dû à une incohérence chez Parfit, celui-ci ne considérant pas les conséquences de ce qu'il affirme quant à l'identité animale, ou alors il y a quelque chose qui nous échappe dans sa position. Cela dit, Parfit lui-même jugera que sa seconde

position est une pauvre tentative de colmater les trous ouverts par la critique animaliste parce qu'elle affaiblit grandement la position lockéenne. Sa troisième position est plus solide à cet égard. À notre avis par contre, moyennant qu'on soit prêt à accepter la dépendance complète de la personne sur le même animal pour sa survie, une position à double critère permettrait de combiner les apports principaux du lockéanisme et de l'animalisme, et d'unifier ainsi notre compréhension de nous-mêmes. Nous argumenterons en faveur d'une telle position « compatibiliste » dans le troisième chapitre (3.2).

*

Passons en revue ce que nous avons étudié dans cette section. D'abord, Parfit reconnaît la force incontournable des arguments animalistes et leur donne raison ; en admettant qu'il est plus plausible d'affirmer que nous sommes numériquement identiques à des animaux humains et que les personnes humaines ne sont qu'une manière de les saisir. Par contre, il tient à traiter les personnes comme des entités conceptuellement distinctes des animaux, ayant leur propre critère d'identité psychologique. De plus, il s'oppose à la thèse de l'animalisme classique selon laquelle notre critère d'identité animale n'inclut pas notre cerveau, parce que cela impliquerait que la personne pourrait survivre séparément de nous et réintroduirait ainsi le problème de dédoublement des penseurs. Par conséquent, il propose que le cerveau serait la partie essentielle de l'organisme, de sorte que personne et animal ne puissent survivre séparément. Nous avons expliqué que pour être acceptable, une telle position devrait montrer que le cerveau remplit des fonctions vitales essentielles que le reste du corps ne pourrait satisfaire, de manière à justifier que nous serions notre cerveau plutôt que notre corps décérébré. Quant au critère de notre identité animale, nous avons montré que Parfit semble refuser les deux options possibles auxquelles mène sa position, indiquant qu'il n'est pas prêt à l'assumer pleinement.

2.3 – *Parfit III* : la personne lockéenne comme partie de l'animal

Voyons à présent la troisième et dernière position de Parfit sur ce que nous sommes. Elle fournit une deuxième réponse aux arguments animalistes, que Parfit adopte

près de vingt ans après celle que nous venons d'examiner. Cette fois-ci, il n'est plus question pour lui de trouver un compromis qui concéderait presque tout aux animalistes tout en sauvant autant que possible sa position initiale, mais de formuler une position qui sache faire face aux objections animalistes tout en parvenant à maintenir que nous sommes numériquement identiques à des personnes lockéennes. Parfit renoue donc avec un lockéanisme plus classique, se réappropriant la distinction lockéenne et niant que nous sommes des animaux.

Pour mieux saisir sa nouvelle position, rappelons le problème principal que l'animalisme identifiait dans le lockéanisme. Si l'animal et la personne sont numériquement distincts, comme le propose le lockéanisme, et que tous deux pensent, cela entraîne une prolifération indue des penseurs dans notre ontologie – c'est l'argument de l'animal pensant (voir 2.1.2). L'animalisme résolvait ce problème en soutenant que la personne n'était pas un deuxième penseur distinct de l'animal, mais bien cet animal lui-même, saisi durant la période où il possède des propriétés mentales particulières ainsi que le statut moral qui leur est associé. La personne pourrait certes être considérée comme une entité pensante, mais seulement en vertu du fait que l'animal qu'elle est pense. Autrement dit, la position animaliste implique que si la personne pense, elle ne le fait que de façon dérivée, héritant ses facultés de l'animal qui les possède en propre. L'apparent dédoublement des penseurs serait ainsi éliminé. C'est là l'argument principal invoqué pour privilégier l'animalisme.

Pour trancher en faveur du lockéanisme plutôt que de l'animalisme, il faudrait pouvoir résoudre le problème de la prolifération des penseurs dans l'autre sens, en soutenant que c'est l'animal qui pense de manière dérivée en vertu des facultés de la personne, ou encore qu'il ne pense pas du tout. Mark Johnston note que certains constitutionnalistes pourraient être tentés de soutenir que c'est la personne qui pense, justement parce que l'animal entre dans *sa* constitution. Mais c'est une vaine entreprise, favorisant ultimement l'animalisme :

« [I]f I had to pick which of the two things I am identical with, the person or the animal, a good rule would be: Pick the thing which is *non-derivatively* the subject of mental acts. And on the present proposal with its appeal to “inheritance” it is the animal, and not the person, that is non-derivatively the subject of my mental acts. » (Johnston 2007 : 50)

Parfit nomme cette règle que formule Johnston le *Own-Thinker Principle* (Parfit 2012 : 18), que nous allons traduire ici par *principe du penseur non-dérivatif*, et qu'Olson attribue à Roderick Chisholm (Olson 2015c ; Chisholm 1976 : 104). Selon ce principe, si la personne hérite ses facultés mentales de l'animal, elle ne peut être le penseur autonome, et ne pourrait donc être *nous*. Attribuer la pensée à la personne à l'exclusion de l'animal et donner ainsi victoire au lockéanisme requiert donc un argument plus élaboré.

Dans cette section, nous allons d'abord voir comment une position constitutionnaliste, celle de Shoemaker, évite de multiplier le nombre de penseurs en attribuant la faculté de penser exclusivement à la personne (2.3.1). Nier à l'animal l'usage de son cerveau est par contre une stratégie contestable, de sorte que Parfit construit une position différente, non-constitutionnaliste, en réponse au problème. Il s'appuie, pour ce faire, sur ce qu'Olson reconnaît être une objection lourde de conséquence pour la position animaliste : le *problème des parties pensantes* (*The Thinking Parts Problem*), que nous allons voir au préalable (2.3.2). Il nous sera alors possible d'examiner la position finale de Parfit, selon laquelle nous sommes des parties pensantes d'animaux plutôt que les animaux eux-mêmes (2.3.3).

2.3.1 – L'animal non-pensant de Shoemaker

Examinons d'abord la réponse de Shoemaker au problème de la prolifération des penseurs. Comme nous l'avons mentionné plus tôt (2.1), Shoemaker est un constitutionnaliste, concevant que tout être humain est constitué de deux entités numériquement distinctes, une personne et un animal. Afin d'éviter de multiplier le nombre d'entités pensantes comme l'implique un constitutionnalisme classique, Shoemaker développe un argument fonctionnaliste qui justifierait qu'on refuse d'attribuer des facultés mentales aux animaux. En un mot, seule une entité ayant des conditions de persistance psychologiques pourrait avoir des propriétés mentales – ce qui exclut les corps, les cerveaux, et les animaux, dont les critères d'identité sont, tel que son fonctionnalisme l'implique, physiques ou biologiques.

Shoemaker justifie sa position ainsi. Dire qu'un sujet a des propriétés mentales, c'est dire qu'il a des pouvoirs causaux qui produisent des états mentaux successeurs appartenant nécessairement à ce même sujet :

« [T]he functional role of a mental state will consist in part in its generating appropriate successor states, and affecting future behaviour, usually in conjunction with other mental states. And those successor states, and that behaviour, will be on the part of the possessor of the mental state. »
(Shoemaker 1999b : 299)

Ainsi, d'après le fonctionnalisme du mental, le fait qu'une entité possède des propriétés mentales a des implications pour la carrière future de cette entité (Shoemaker 1999b : 298). L'entité qui possède les intentions d'agir découlant de croyances préexistantes *est* l'entité qui avait ces croyances. Autrement dit, le rôle fonctionnel des états mentaux implique que le sujet à qui appartient une série d'expériences causalement interreliées survit en tant que le sujet dont les expériences poursuivent cette même série et en découlent causalement. Le critère de persistance de toute entité pensante serait donc nécessairement psychologique.

Un tel fonctionnalisme s'applique également aux propriétés physiques et biologiques selon Shoemaker. Une entité ayant des états biologiques tel un animal survit alors en tant que l'entité ayant ses états biologiques successeurs. Par exemple, un animal humain dont l'estomac digère un aliment serait le même animal qui poursuit et termine cette digestion. Or, puisque la vie mentale pourrait être séparée de l'organisme qui remplit de telles fonctions biologiques, par la mise en cuve de son cerveau par exemple, personne et animal ont donc des conditions de persistance différentes et sont numériquement distincts. Le fonctionnalisme justifierait ainsi la distinction lockéenne. Cette vie mentale n'aurait alors jamais appartenu à l'animal, mais serait le propre de la personne.

Il serait certes possible de défendre que le cerveau supportant la vie mentale de la personne *est* l'animal, comme l'a fait le second Parfit, mais ce n'est pas la voie qu'emprunte Shoemaker :

« [T]he person is not the cerebrum, and so the physical properties that realize the person's mental properties should be properties of the person, not properties of the cerebrum. » (Shoemaker 1999b : 299)

Selon lui, les propriétés physiques qui réalisent les états mentaux de la personne sont exclusivement des propriétés de cette personne ; l'animal ne les a jamais eues. Par suite, puisque la personne a des conditions de persistance psychologiques, les propriétés physiques qui réalisent ses états mentaux présents et futurs doivent, en vertu de leur rôle fonctionnel, avoir ces mêmes conditions de persistance :

« [W]here the mental property is apt to produce certain successor mental states, the realizer property must be apt to produce certain successor physical states, these being realizations of the successor mental states. And, as before, these successor states must belong to the same subject as that produce them. » (Shoemaker 1999b : 300)

Les conditions de persistance de la personne *et* de ses réalisateurs physiques seraient donc *purement* psychologiques. La continuité psychologique inclut donc la continuité de ses réalisateurs physiques sans nécessiter de clause additionnelle telle la continuité de suffisamment du cerveau, contrairement à ce que Parfit défend depuis 1987.

Shoemaker considère même la possibilité que la personne ne soit pas rattachée à des réalisateurs physiques particuliers. S'il était possible de faire une copie de l'état des réalisateurs physiques d'une personne (*Brain State Transfer*), il serait concevable que l'identité personnelle soit préservée par une instanciation matérielle de cet état (Shoemaker 1984 : 108 ; 1999b : 293). Il admet par contre que les « personnes » dont la survie est assurée par un tel mécanisme puissent plutôt être des entités différentes des personnes tel qu'on les conçoit normalement, le dispositif de transfert de leurs états mentaux faisant alors partie intégrante de leur identité et de leur nature (Shoemaker 1984 : 111).

La position de Shoemaker se présente donc comme une version du néo-lockéanisme très proche de Locke, mais matérialiste. Elle a le mérite de parvenir à réconcilier l'idée que nous sommes de pures consciences avec celle que nous sommes des entités réelles faisant partie du monde naturel, contrairement aux positions cartésiennes, lockéennes et parfitiennes. On peut même voir dans le fonctionnalisme de Shoemaker une explication du *principe d'exclusion* lockéen : deux entités seraient de type différents et pourraient coïncider si et seulement si elles possèdent des propriétés de nature différente, dont les rôles fonctionnels déterminent des règles de génération d'états successeurs distinctes pour chaque entité, leur conférant ainsi des conditions de persistance différentes. L'animal, dont les fonctions biologiques déterminent qu'il persiste en tant que l'entité qui a ses états biologiques successeurs, ne pourrait donc penser parce qu'il n'a pas les bonnes conditions de persistance pour le faire. Ce serait aussi le cas de tout organisme.

Malgré l'argumentation serrée que met en œuvre Shoemaker, il y a de fortes

objections qui peuvent être dirigées contre sa position, notamment celles d'origine évolutionniste. La sélection naturelle de traits menant au développement de facultés mentales n'a-t-elle pas lieu car elle favorise la survie de l'*animal* ? Pourquoi des animaux tels que nous-mêmes dépenseraient-ils tant d'énergie pour alimenter un cerveau s'ils n'avaient pas accès aux facultés mentales qu'il supporte ? Olson abonde dans le même sens : « What stops a typical human animal from using its brain to think? Isn't that what that organ is *for*? » (Olson 2003a : 15) La troisième position de Parfit peut être comprise comme une version du néo-lockéanisme qui tente d'échapper à de telles difficultés, quoiqu'elle ait son propre lot de problèmes, comme on va le voir plus loin (2.3.3).

2.3.2 – Le problème des parties pensantes

Tournons-nous maintenant vers une objection majeure à l'animalisme, d'où est issue la troisième position de Parfit : le *problème des parties pensantes*. Celui-ci peut être résumé ainsi. Si la pensée peut être tout autant attribuée à une partie de l'animal qu'à l'animal lui-même, alors l'idée que *nous*, le penseur non-dérivatif, sommes l'animal, n'est pas une thèse aussi évidente que ne le croient les animalistes, même lorsqu'on écarte la personne des candidats possibles. Nous pourrions n'être qu'une partie de l'animal – son cerveau, ses hémisphères, ou encore, tel que Parfit le défend, la personne, si elle est reconceptualisée comme une partie de l'animal. Nous allons nous en tenir à l'objection ici ; la position de Parfit sera l'objet de la prochaine sous-section (2.3.3).

Barry Dainton, dans sa critique de l'ouvrage d'Olson *The Human Animal*, est le premier sinon un des premiers à formuler cette objection, et préfigure le troisième Parfit en suggérant que la personne pourrait être une partie de l'animal :

« Yes, animals have mental capacities, but strictly speaking, the whole animal does not possess these capacities, only a part of the animal possesses them; so in saying that "Animals have mental capacities", what we are really saying—or would say on reflection—is that "Animals have mental capacities by possessing a part which has these capacities". Since (proper) parts are numerically distinct from their wholes, if we associate persons with the relevant parts, the claim that we are animals is undermined. » (Dainton 1998 : 681)

Cette objection est hautement problématique pour l'animalisme. L'argument de l'animal pensant proposait, contre le lockéanisme, un candidat autre que la personne au titre du

penseur autonome que nous sommes. Il nous fallait alors choisir entre deux candidats, et comme les animaux sont des êtres matériels, il semblait plus plausible de réduire la personne et de nous identifier à l'animal. Mais si une partie de l'animal est également un candidat admissible, nous nous retrouvons alors devant une nouvelle version du problème de la multiplication des penseurs, opposant cette fois-ci le tout à la partie.

Disposons-nous d'un argument qui nous permettrait alors de défendre que nous sommes le tout plutôt que la partie pensante ? Au contraire, le *principe du penseur non-dérivatif*, qui nous somme de nous identifier à l'entité qui pense par elle-même plutôt qu'à celle qui en hérite les facultés, tranche en faveur de la partie plutôt que du tout. Autrement dit, si une partie de l'animal pense, alors *nous*, le penseur non-dérivatif, ne serions pas des animaux mais plutôt cette partie pensante. Et si la personne était cette partie pensante, alors nous serions numériquement identique à elle. L'argument de l'animal pensant semble ainsi avoir perdu toute la force qui nous inclinait à préférer l'animalisme au lockéanisme.

Or, la partie de l'animal qui se présente intuitivement à l'esprit comme penseur est le cerveau, conduisant à la version du problème de la multiplication des penseurs qu'Olson nomme le *problème du cerveau pensant* (*The Thinking Brain Problem*) (Olson 2007 : 78-84). Si mon cerveau pense, et qu'il n'y a qu'un seul penseur habitant mon corps, alors, à strictement parler, si je suis l'animal, je ne pense pas ; mon cerveau pense pour moi. Une telle difficulté nous mène à dire que, suivant le *principe du penseur non-dérivatif*, nous sommes notre cerveau plutôt que l'animal – quoiqu'il y ait moyen de s'extirper de cette objection en défendant l'idée que le cerveau *est* l'animal, comme nous l'avons fait dans la section précédente (2.2.3).

Cependant, Olson montre que ce problème est généralisable, de sorte que la réponse qui identifie l'organisme au cerveau s'avère franchement insuffisante. Si le cerveau apparaît de toute évidence comme premier candidat au rôle de la partie pensante de l'animal, il n'est pas le seul. En effet, il y a une multitude d'autres objets que nous pourrions sélectionner comme partie pensante : les hémisphères cérébraux, la tête, la moitié supérieure du corps, et le cerveau plus une cellule, par exemple, seraient tout autant capables d'engendrer nos pensées. Toute subdivision de l'animal qui inclut les hémisphères – ou une plus grande part du cerveau si les hémisphères sont insuffisants –

dispose des facultés mentales requises. Il y aurait donc une infinité de candidats au titre du penseur autonome chez un même animal. Mais, à moins qu'on admette cette absurde prolifération d'entités dans notre ontologie, il faudrait pouvoir en identifier une seule comme penseur non-dérivatif. C'est ce qu'Olson nomme le *problème des parties pensantes* (Olson 2007 : 215-9).

Remarquons que la multiplication des penseurs vient accompagnée de problèmes corollaires – comme c'était le cas avec le problème de l'animal pensant. Chacun de ces penseurs, en vertu de ses facultés de pensée rationnelle et de conscience de soi, se qualifie alors comme une personne selon la définition lockéenne. Avec la prolifération des penseurs vient la prolifération des personnes. De plus, comme chacun de ces penseurs coïncide avec les autres et partage la même matière, et donc les mêmes pensées, ils ont tous les mêmes raisons de croire qu'ils sont le penseur non-dérivatif, quoiqu'à l'exception d'un seul d'entre eux, ils se trompent tous. Nous n'aurions donc pas de moyen de déterminer de façon non-arbitraire lequel d'entre eux nous sommes ; c'est le *problème épistémique* qu'identifiait plus tôt Olson.

Bref, l'argument de l'animal pensant, qui introduisait un second penseur pour faire face à la personne, se trouve ainsi généralisé et retourné contre l'animalisme. Olson juge d'ailleurs que c'est une des plus fortes objections à sa position :

« [A]nimalism has the virtue of avoiding the thinking-animal problem. But if brains and heads also think our thoughts, it is hard to see how we could have any reason to think that we are animals either. » (Olson 2007 : 216)

Alors que pour résoudre le problème de l'animal pensant en faveur de l'animalisme, il fallait seulement montrer que l'animal était un candidat plus plausible que la personne au titre de penseur autonome ; pour résoudre le problème des parties pensantes en faveur de l'animal, cela requiert d'éliminer tout autre candidat concevable, c'est-à-dire de montrer en quoi l'animal est la seule option crédible parmi une infinité. La demande faite à l'animalisme est désormais considérablement plus élevée.

Il est certes possible de généraliser la solution de van Inwagen au problème du dédoublement des penseurs, que Parfit empruntait pour appuyer sa seconde position, afin de l'appliquer à la prolifération sans fin des êtres pensants. Comme nous l'avons vu, il n'y a dans l'ontologie de van Inwagen que des êtres vivants et des particules simples. Les parties d'êtres vivants n'existent pas réellement ; elles ne sont que des subdivisions

conceptuelles de l'entité réelle. Prises individuellement, les parties non-détachées de l'animal ne sont qu'un ensemble de particules arrangées en une certaine forme. Si une subdivision de l'animal était détachée du reste et demeurerait capable de préserver l'existence du sujet pensant, c'est qu'elle *serait* cet animal, mais mutilé. Par suite, s'il n'y a pas de cerveau, d'hémisphères, ni de têtes, nous n'avons pas à nous soucier que nous pourrions être ces parties d'animal (Olson 2007 : 218).

Selon Olson, si ce n'est pas la seule solution possible pour refuser l'éligibilité des candidats rivaux au titre de penseur, c'est du moins la plus simple et la meilleure (Olson 2007 : 228). Par contre, elle implique des engagements ontologiques forts que bien des philosophes ne seraient pas prêts à accepter. Notamment, elle requiert l'adhésion à une ontologie minimaliste, réductionniste à propos de tout objet inorganique ainsi que des objets organiques qui ne sont pas des êtres vivants, tels les organes et les tissus. Elle découle d'une théorie de la composition qu'Olson nomme le *minimalisme biologique*, empruntée à van Inwagen, selon laquelle des particules ne composent une entité réelle que lorsque leur activité constitue une vie (Olson 2007 : 226). Une telle position exige un support solide avant d'être adoptée, et s'expose d'ailleurs à au moins une objection majeure. Joshua Watson avance que si l'adoption du minimalisme biologique réglerait bien des problèmes pour l'animalisme, toute défense d'un tel minimalisme présuppose à son tour l'animalisme, ou repousse le problème de la prolifération des penseurs à un autre niveau (Watson 2016 : 327). Nous n'entrerons pas dans ses arguments ici.

Bref, si les parties de l'animal sont également des candidats au titre du penseur non-dérivatif que nous sommes, alors la thèse selon laquelle nous sommes des animaux n'est pas aussi évidente qu'elle n'y paraît. Si l'animal ne pense qu'en vertu de la partie qui a la faculté de penser, nous serions alors cette partie suivant le *principe du penseur non-dérivatif*. Mais *quelle* partie d'animal serions-nous alors ?

2.3.3 – La personne comme partie pensante incarnée

Voyons maintenant comment Parfit s'appuie sur le *problème des parties pensantes* pour construire sa nouvelle position. Il croit avoir une solution lockéenne « évidente » (2012 : 14) à ce problème, qui évite les difficultés auxquelles l'animalisme se trouve confronté, et qui résout également le problème original de l'animal pensant en faveur du lockéanisme, contrairement à sa seconde position :

« According to another, better view, we are not animals, or human beings. We are what McMahan calls the conscious, thinking, and controlling *parts* of human beings. We can call this the *Embodied Part View*. The Thinking Parts Problem has a thinking parts solution. » (Parfit 2012 : 14)

Nous serions alors une partie de l'animal, celle qui pense par elle-même et contrôle le corps dans lequel elle est incarnée. Selon Parfit et conformément à l'approche psychologique lockéenne, la personne serait cette partie. Nous allons examiner comment il justifie cette affirmation.

Notons avant tout que cette position aurait l'avantage d'éviter l'objection que nous avons soulevée contre le constitutionnalisme de Shoemaker. En avançant que seule la personne possède les propriétés mentales et ses réalisateurs physiques, Shoemaker se trouvait à dire que l'animal ne pouvait utiliser son cerveau pour penser. Mais selon cette nouvelle position de Parfit, l'animal utilise bien une partie de lui-même pour penser, tout comme il digère ses aliments en ayant une partie, son estomac, qui le fait pour lui (Parfit 2012 : 14). Le fait qu'une partie remplisse les fonctions normalement attribuées au tout n'ôte pas à l'animal les propriétés et fonctions de ces parties ; il les a justement en vertu du fait que ses parties les ont, il en hérite. Par suite, il n'y a pas de dédoublement d'entités pensantes ou digérantes ; à strictement parler, seule la partie pensante pense en propre, et seul l'estomac digère. L'animal n'est qu'un penseur et un digéreur dérivatif.

Afin d'expliquer pourquoi la personne serait la partie pensante de l'animal, Parfit passe en revue les options. Quoique le cerveau ou les hémisphères soient des candidats recevables, Parfit les rejette en raison d'arguments mis de l'avant par Olson et Johnston : si nous étions littéralement des cerveaux (et non pas des animaux dont la partie essentielle est l'organe cérébral), cela impliquerait que nous ne serions qu'une masse de tissus d'un kilo et demi emprisonnée dans notre tête depuis notre naissance, une absurdité selon eux (Johnston 2007 : 55 ; Olson 2015c : 10). Parfit se défait donc de l'idée que le

cerveau est l'organe qui pense et contrôle le reste du corps – idée qu'il défendait dans sa seconde position. Serait-ce alors une autre subdivision du corps qui penserait pour nous ?

Parfit répond par la négative. Quoiqu'il n'invoque pas directement la raison suivante pour écarter les parties du corps animal comme penseurs potentiels, toute partie du corps qui pourrait être proposée, tant qu'elle contient les hémisphères cérébraux, se trouverait confrontée à une infinité de rivaux tout autant éligibles, comme nous l'avons vu. Afin d'éviter la prolifération des penseurs, il faudrait pouvoir dire pourquoi la partie du corps choisie pourrait penser alors que l'ajout ou le retranchement d'une seule cellule, qui en ferait une partie différente, lui enlèverait cette faculté. Or, depuis ses premiers écrits, Parfit défend l'idée que nous ne sommes pas notre corps, ni notre cerveau, mais *ce qui* a un corps et un cerveau, conformément à notre usage habituel de la langue. Nous sommes des êtres distincts des parties constituant notre corps pour la simple raison que nous nous concevons ainsi. Toute subdivision du corps de l'animal est ainsi rayée de la liste interminable de candidats au titre de partie pensante, coupant ainsi court à la prolifération des penseurs.

Mais si toute partie du corps est exclue, quelle serait alors la partie pensante que nous sommes ? La première option que Parfit considère est la position de Jeff McMahan. Ce dernier propose que nous serions des *esprits (minds)* incarnés, parce que cette notion est équivalente à celle de la « même conscience » mise de l'avant par Locke, et parce qu'elle saisit ce qui importe pour nous, contrairement à l'identité personnelle (McMahan 2002 : 67). En effet, comme nous l'avons vu dans le premier chapitre, Parfit a montré à l'aide d'expériences de pensée que même si notre identité personnelle n'est pas maintenue lors de la duplication de notre continuité psychologique – puisque l'identité est une relation qui requiert l'unicité – ce qui importe pour nous, notre vie mentale, est tout de même préservée, et même doublement !

Parfit rejette cependant cette version de la *Embodied Part View*, parce qu'elle partage certains problèmes avec la version qui nous identifie à notre cerveau ou nos hémisphères. Il rapporte l'argument de William Carter, qui affirme que puisque nous avons des bras et des jambes, mais pas les esprits, nous ne pouvons alors être identifiés à ces derniers (Parfit 2012 : 16). Parfit évoque également l'argument de Johnston selon lequel la logique de notre grammaire nous indique que les sujets auxquels il est correct

d'appliquer des prédicats mentaux sont l'animal et la personne, plutôt que des parties corporelles ou des esprits (Johnston 2007 : 54). Selon Johnston, si l'on ne se tient pas à cette logique de notre grammaire, qui n'a rien d'accidentel, nous serons vite perdus. C'est pourquoi Parfit propose une version de la *Embodied Part View* selon laquelle la partie pensante de l'animal est une personne lockéenne :

« [H]uman animals think by having a conscious thinking part which is a person in the Lockean sense. We can call this the *Embodied Person View*. This, I believe, is the best version of the Embodied Part View. » (Parfit 2012 : 17)

Parfit explique que, contrairement aux autres versions de la *Embodied Part View* selon lesquelles nous serions des cerveaux ou des esprits, il est tout à fait cohérent d'attribuer les propriétés physiques du corps à la personne, tel le fait d'avoir des bras. En tant que personnes *incarnées*, nous sommes intimement liés au corps que nous incarnons, et pouvons faire un usage élargi du « je » pour nous approprier notre corps, comme nous disons être mouillés lorsque nos habits le sont (Parfit 2012 : 20). Les propriétés que nous nous attribuons, physiques et mentales, peuvent donc toutes être attribuées à la personne, de sorte qu'il n'y a rien qui nous empêche d'affirmer que nous sommes numériquement identiques à cette personne.

Par suite, nous ne serions pas essentiellement des animaux. Nous pourrions survivre à la mort de l'animal dans lequel nous nous incarnons pour autant que la personne survive. Conformément à la *Transplant Intuition*, nous, la personne, irions là où iraient nos hémisphères, parce nous sommes « the part whose physical basis is the cerebrum. » (Parfit 2012 : 18) Et en se ré-incarnant dans un nouveau corps, nous gagnerions alors ses propriétés physiques.

Si l'on s'en tient à Parfit, cette position aurait réponse à toutes les objections soulevées jusqu'ici. C'est d'après lui la seule position qui évite le problème des parties pensantes, et qui prévient la prolifération des penseurs et des personnes :

« Since the animal thinks only by having a part that thinks, there are not too many thinkers here. And since the animal is a person only in the derivative sense of having a Lockean person as a part, there are not too many persons here. » (Parfit 2012 : 24)

Ensuite, malgré que selon cette conception la personne et l'animal seraient des entités numériquement distinctes qui partageraient leurs pensées, l'une les pensant directement,

l'autre indirectement, cette position nous permettrait de distinguer laquelle nous sommes sans ambiguïté et se soustrairait ainsi au *problème épistémique* qu'a soulevé Olson. En effet, puisque c'est seulement la personne qui pense à strictement parler, l'animal est alors cognitivement dépendant d'elle :

« Just as the animal thinks some thought only by having a part, the Lockean person, that does the thinking, the animal can know something only by having a part, the person, that knows this thing. » (Parfit 2012 : 23)

Ainsi, l'animal ne peut croire et savoir que ce que la personne croit et sait, et ne pourrait donc se tromper sur le référent du « je » que si la personne se trompe. Mais si nous précisons nos pronoms comme le suggère Parfit, en distinguant le « je-interne » qui réfère à la personne du « je-externe » qui réfère à l'animal, la personne ne pourrait alors se tromper si elle en fait un usage correct. Lorsque nous nous disons « je suis une personne », si le « je » est précisé au sens du « je-interne », l'énoncé demeure vrai qu'on l'attribue à la personne ou à l'animal. L'ambiguïté serait ainsi levée. Parfit considère donc que la capacité qu'a la *Embodied Person View* à répondre aux objections adressées au lockéanisme classique ainsi qu'à éviter les difficultés auxquelles font face les animalistes fournit autant de raisons pour la préférer à ces positions.

En somme, Parfit se saisit du problème des parties pensantes comme d'une occasion pour revenir à une position qui réhabilite la distinction lockéenne entre personne et animal. La personne serait une entité numériquement distincte de l'animal, parce qu'elle serait une *partie* de l'animal. Et nous serions cette personne plutôt que l'animal, puisque nous sommes un penseur non-dérivatif et que le *principe du penseur non-dérivatif* tranche en faveur de la partie pensante plutôt que du tout qui hérite de ses facultés. Il faudrait certes justifier que c'est bien la personne qui pense par elle-même, et Parfit n'offre ici comme justification que le fait qu'elle semble éviter les problèmes auxquels font face les autres candidats, l'animal inclus. Par contre, nous allons voir dans le prochain chapitre qu'une telle justification entraînerait plus de problèmes qu'elle ne prétend en résoudre (3.1.2).

*

Jetons un regard d'ensemble sur l'évolution de la position de Parfit suite à l'entrée en scène de l'animalisme dans le débat concernant notre nature ontologique et notre identité. Face aux arguments animalistes, Parfit a été contraint de revoir sa position afin

de rattacher notre existence à celle de l'animal. La *relation de constitution* qu'il défendait entre la personne et le corps ainsi que le cerveau a été exposée comme hautement problématique lorsque appliquée à l'animal, parce qu'elle impliquait qu'il y avait deux penseurs qui coïncident dans le même espace. Il fallait donc une autre relation pour expliquer notre lien à l'animal. La deuxième position de Parfit utilisait la *relation d'identité* pour le faire, se rangeant du côté des animalistes en affirmant que nous *sommes* des animaux. Cette position compromettait cependant son lockéanisme et avait des implications ontologiques qu'il n'acceptait pas. Sa troisième position établit notre rapport à l'animal autrement, à l'aide d'une *relation méréologique* : nous sommes la partie du tout qu'est l'animal. Une telle position lui permet de maintenir que nous sommes numériquement identiques à des personnes, et de revenir à une position lockéenne plus classique qui correspond davantage à ses convictions, sans perdre de vue notre origine animale.

Cependant, cette relation méréologique est-elle autre chose qu'une relation de constitution inversée ? Plutôt que d'affirmer que la personne est constituée d'un animal, on se trouve à affirmer que l'animal est constitué d'une personne et d'une multitude d'autres parties. Mais si la relation de constitution permet, comme l'a défendu Parfit, de réduire ontologiquement une entité à ses constituants, cette manière de voir les choses revient-elle alors à nier l'existence réelle de l'animal ? Ce serait donc là une piètre manière d'expliquer en quoi nous sommes rattachés à un animal. Pire, si la personne était une partie d'animal, mais que les animaux n'avaient pas de réelle existence, la personne ne serait alors qu'une partie de rien du tout. Nous élaborerons sur cette objection et en développerons d'autres dans le chapitre suivant (3.1).

CHAPITRE III – UNE ALTERNATIVE ANIMALISTE

Dans les deux premiers chapitres de ce mémoire, nous avons fait le tour de l'évolution de la pensée de Parfit à propos de notre identité et de la sorte d'entité que nous sommes. Nous avons identifié trois moments distincts chez lui : son constitutionnalisme initial (ou réductionnisme constitutif, tel qu'il le nomme), son rapprochement avec l'animalisme, et finalement sa position selon laquelle la personne serait une partie de l'animal. Ce faisant, nous avons présenté les raisons motivant l'adoption de chacune de ces positions, en particulier les critiques animalistes contre le lockéanisme ainsi que la position animaliste classique que défend Olson.

Cependant, la position finale de Parfit ainsi que celle d'Olson nous semblent toutes deux insatisfaisantes. Elles ne rendent pas pleinement compte de ce que nous sommes, tronquant notre compréhension de l'animal humain comme *penseur*, *sujet d'expérience* et *agent*. En effet, la position de Parfit n'attribue l'activité mentale qu'à une partie de l'animal plutôt qu'à l'animal lui-même ; alors que la position d'Olson implique que la faculté de penser et d'avoir des expériences n'est pas essentielle à la nature humaine – ni à celle des autres animaux conscients d'ailleurs.

Dans ce troisième chapitre, nous allons d'abord présenter deux objections majeures à la position finale de Parfit (3.1). Nous allons ensuite esquisser une position mitoyenne qu'on pourrait dire « compatibiliste », qui s'apparente à celle de nous avons développée en présentant la deuxième position de Parfit et qui marierait les apports de l'animalisme et du lockéanisme, quoiqu'elle nécessite certains engagements ontologiques plausibles mais non-établis (3.2).

3.1 – Critique de la position finale de Parfit

Dans cette section, nous allons poser un regard critique sur la position de Parfit selon laquelle nous sommes une partie de l'animal, la personne, afin d'en faire ressortir les faiblesses et les incohérences. D'abord, nous allons nous attarder à évaluer comment

cette position s'agence avec son réductionnisme de la personne (3.1.1). Nous avons déjà soulevé cette critique : est-il cohérent d'affirmer qu'une entité qui n'a pas d'existence réelle serait capable de penser par elle-même ? Ensuite, nous allons examiner le *principe du penseur non-dérivatif*, et s'il implique bien, comme Parfit le croit, que ce serait la personne qui pense par elle-même (3.1.2). Nous allons également considérer les conséquences de l'application d'un tel principe, une fois généralisé, à notre compréhension de l'animal.

3.1.1 – Le réductionnisme de la personne et le penseur non-dérivatif

Commençons par évaluer la position finale de Parfit à la lumière de son réductionnisme ontologique de la personne. Nous allons développer un argument selon lequel si la personne n'est qu'une entité de nature conceptuelle, qu'un sujet grammatical tel que l'implique le réductionnisme de Parfit, elle ne serait tout simplement pas la sorte d'entité capable de penser par elle-même. Elle ne pourrait donc être la partie proprement pensante de l'animal.

Rappelons avant tout que la position finale de Parfit, qui nous identifie à la partie pensante de l'animal, c'est-à-dire la personne selon lui, se veut une réponse à un problème fort embêtant pour l'animalisme : le *problème des parties pensantes*. En effet, toute subdivision du corps qui comprend le cerveau se qualifie comme penseur, de sorte que l'animal se trouve opposé à une infinité de candidats rivaux à ce titre. Or, puisque nous sommes l'entité qui pense par elle-même plutôt que de tenir ses pensées d'une autre entité qui penserait pour elle, suivant le *principe du penseur non-dérivatif*, l'animalisme est alors placé devant la tâche considérable d'expliquer pourquoi tous ces penseurs ne pensent pas par eux-mêmes, mais que seul l'animal pense en propre. La solution de Parfit évite ce problème en sélectionnant comme penseur une partie de l'animal qui n'est pas une subdivision de son corps, mais une entité de nature conceptuelle : la personne.

Cette solution, construite dans le seul but de répondre au *problème des parties pensantes*, peut paraître arbitraire. Qu'est-ce qui nous permet d'affirmer que la personne est effectivement une partie de l'animal ? D'ailleurs, où se trouve-t-elle localisée chez l'animal ? Et qu'est-ce qui justifierait l'idée que c'est elle qui pense en propre, plutôt qu'une autre partie du corps tel le cerveau ?

La position ne prend son sens que comprise à travers la lentille réductionniste de Parfit. Comme nous l'avons expliqué dans le premier chapitre, c'est toujours, selon Parfit, parce que nous nous distinguons usuellement des parties de notre corps que nous, la personne, sommes une entité qui en est distincte. Cependant, cette distinction n'est que conceptuelle : la personne n'a pas d'existence réelle, car son existence se réduit à celle des composantes qui la constituent, en l'occurrence les hémisphères cérébraux et un ensemble d'événements mentaux interreliés. La personne n'est donc une partie supplémentaire de l'animal que parce qu'on ne se considère pas littéralement comme une partie de notre corps. Elle n'est pas physiquement localisée, mais habite notre schème conceptuel. C'est elle qui pense en propre, parce que Locke la définit comme entité essentiellement pensante et que nous sommes ce qui pense par soi-même. Bref, la solution de Parfit est purement grammaticale et stipulative.

Cependant, le réductionnisme ontologique de la personne et l'affirmation que celle-ci est l'entité qui pense de façon autonome sont à notre avis des propositions incompatibles. En effet, puisque la personne n'a qu'une existence conceptuelle, il semble incohérent de lui attribuer la faculté de penser par elle-même, et ce à la place d'entités matérielles comme les animaux, ou encore les cerveaux (si nous ne sommes pas réductionnistes à propos des parties d'animaux comme van Inwagen et Olson). Affirmer que les concepts peuvent penser est tout simplement absurde.

Asseyons cette affirmation sur un argument plus élaboré. À notre avis, s'il faut choisir une entité à laquelle nous sommes numériquement identiques, et qui serait capable de penser de façon autonome, une bonne règle serait de choisir une entité ayant une existence réelle, qui soit capable de pratiquer une activité produisant des effets dans le réel. Autrement dit, nous nous appuyons sur le principe provisoire suivant, qu'il paraît plausible d'adopter :

Principe d'effectivité : Seules des entités réelles ont des pouvoirs causaux et peuvent produire des effets réels.

Nous supposons ici que la pensée est une activité ayant des effets réels, concrets et observables, plutôt que de n'être qu'un épiphénomène. Parfit fait de même d'ailleurs lorsqu'il avance que la partie pensante de l'animal est la partie qui contrôle son corps (Parfit 2012 : 14). Si cette supposition et le *principe d'effectivité* ne sont pas faux, cela

signifie que seule une entité réelle pourrait penser par elle-même. Une entité n'ayant qu'une existence conceptuelle, c'est-à-dire dont l'existence se réduit à l'existence d'autres entités comme c'est le cas de la personne dans la conception de Parfit, n'aurait pas de pouvoir causal ; elle ne pourrait alors produire d'effets réels que de manière dérivée, par l'entremise d'une entité réelle qui produit ces effets directement.

Par suite, pour pouvoir affirmer de manière cohérente que la personne pense en propre tout en maintenant son réductionnisme, Parfit devrait défendre l'idée que des entités irréelles pourraient avoir des pouvoirs causaux et produire des effets dans le réel directement, sans médiation par une entité réelle. C'est certes une croyance commune que les idées ont un pouvoir causal et qu'elles peuvent influencer l'agir des individus qui y sont exposés ; toutefois, les idées ne sont qu'un moyen que les individus emploient pour influencer le comportement d'autrui, et n'ont pas d'effet indépendamment des individus qui les considèrent et les font entrer dans la détermination de leur agir. Nier le *principe d'effectivité* ne semble pas être une avenue facile ni plausible.

L'alternative qui permettrait à Parfit de conserver sa position de la personne comme partie pensante de l'animal serait de rejeter une des deux propositions que le *principe d'effectivité* expose comme étant incompatibles : soit le réductionnisme ontologique de la personne, soit l'idée que la personne pense de façon autonome. Considérons les branches de ce dilemme tour à tour.

Si l'on opte pour l'abandon du réductionnisme, cela revient à affirmer que la personne est une entité réelle qui pourrait alors avoir des pouvoirs causaux et produire des effets réels, conformément au *principe d'effectivité*. Cela justifierait l'idée que la personne pense par elle-même. Il faudrait par contre expliquer en quoi la personne est une telle entité, ayant une existence réelle et distincte de celle de l'animal, des parties de son corps, et de ses expériences. On pourrait affirmer soit que la personne est une substance immatérielle, adoptant le dualisme et défendant que l'âme est une partie additionnelle de l'animal, soit que la personne est numériquement identique à une partie du corps physique. Si on embrasse le dualisme, on se retrouverait face au problème épineux de l'interaction causale entre entités matérielles et immatérielles. Mais puisque Parfit rejette catégoriquement le dualisme, et qu'il croit que les hémisphères cérébraux sont le support physique de la personne (Parfit 2012 : 18), il devrait alors soutenir que la

personne *est* les hémisphères, plutôt que de considérer qu'ils sont distincts.

Le problème, c'est que nous voilà de retour au réductionnisme, la personne n'étant rien d'autre que ses hémisphères, qui, *eux*, seraient l'entité réelle. Et puisque ce serait l'activité des hémisphères qui produit directement notre pensée, la personne ne penserait qu'en vertu de ceux-ci. Parfit se trouverait alors à affirmer que nous, la partie pensante de l'animal, sommes nos hémisphères plutôt que la personne, suivant les *principes du penseur non-dérivatif et d'effectivité*. Il semble donc qu'à moins d'être dualiste, il est bien difficile de justifier que la personne soit une entité réelle – et encore dans le cas du dualisme c'est loin d'être partie gagnée.

Pour défendre l'idée que c'est la personne qui pense plutôt que ses hémisphères, la meilleure option pour Parfit serait peut-être d'adopter une position telle celle de Shoemaker selon laquelle les propriétés physiques du cerveau qui réalisent l'activité mentale sont des propriétés appartenant exclusivement à la personne. Autrement dit, les hémisphères ne seraient pas une partie de l'animal mais constitueraient la personne, qui, *elle*, serait une réelle partie de l'animal. Il est toutefois douteux qu'un tel tour de passe-passe linguistique fasse toute la différence.

Néanmoins, une telle position se froterait de nouveau au *problème des parties pensantes* : en effet, pourquoi serait-ce les hémisphères qu'on identifie à la personne, plutôt qu'une autre subdivision du corps qui comprend les hémisphères ? Parfit se trouverait alors confronté à la tâche d'éliminer l'infinité de candidats rivaux – une difficulté majeure imputée à l'animalisme – alors que sa position était construite expressément pour l'éviter. La raison d'être de la position finale de Parfit serait remise en question. De plus, Parfit se verrait dans l'impossibilité de nous attribuer les propriétés physiques que nous nous reconnaissons normalement, notamment celles d'avoir des bras et des jambes. En effet, une partie du corps de l'animal ne peut avoir les propriétés physiques des autres parties du corps, sans quoi ce ne serait pas à proprement parler une *partie* du corps. Bref, la *Embodied Person View* de Parfit serait en bien moins bonne posture que l'animalisme, malgré les problèmes auxquels fait face ce dernier.

D'ailleurs, si la personne avait une existence réelle et était la partie proprement pensante de l'animal, Parfit devrait alors se défaire de l'idée que notre vie mentale pourrait être pleinement intelligible indépendamment de personnes ou sujets (voir 1.4).

Rappelons que Parfit justifie cette thèse ainsi : 1) les expériences pourraient, dans un autre schème conceptuel que le nôtre, dépendre conceptuellement d'autre chose que des personnes, comme les corps ou les séries d'expériences ; et 2) les expériences pourraient dépendre ontologiquement des cellules du cerveau dans lesquels elles sont réalisées. Cependant, si, plutôt que des ensembles de cellules, c'est la personne qui est le penseur non-dérivatif, cela implique que nos expériences dépendent ontologiquement des personnes. Omettre alors les personnes de l'explication de notre vie mentale reviendrait à en omettre la cause réelle, ce qui nous empêcherait de la comprendre pleinement.

Revenons au dilemme. Si l'on emprunte la seconde branche, celle de nier que la personne pense par elle-même, cela signifierait que *nous*, le penseur autonome, ne sommes pas numériquement identiques à des personnes. Peu importe si l'on imputait la pensée à l'animal, aux hémisphères cérébraux, ou à une autre partie du corps, la personne ne penserait alors que de manière dérivée, par l'entremise de l'entité réelle qui pense en propre, héritant de ses pensées. Quoiqu'on pourrait maintenir que la personne reste une partie « pensante » de l'animal, comme toute subdivision du corps comprenant le cerveau, elle ne serait pas *la* partie proprement pensante, *nous*. Ainsi, une telle position ne permettrait pas à Parfit de maintenir son lockéanisme, puisque nous ne serions pas identiques à des personnes, et que les personnes ne seraient pas à proprement parler des penseurs.

De plus, la personne se trouvant alors pleinement réduite à l'entité réelle capable de penser par elle-même, la question suivante se poserait alors : pourquoi dire que la personne est une partie supplémentaire de l'animal, si elle n'a aucune effectivité ? Ne serait-ce pas gonfler inutilement notre ontologie ? Quelle raison aurait-on alors de considérer la personne comme une partie de l'animal plutôt que comme un prédicat attribuable au réel penseur, autrement dit comme un *sortal de phase* de ce dernier ? Lorsqu'on nie que la personne pense par elle-même, la *Embodied Person View* de Parfit apparaît comme une construction intellectuelle superflue. Le seul aspect du lockéanisme qu'elle lui permettrait de conserver serait l'idée que la personne est numériquement distincte de l'animal, et ce seulement si nous attribuons la pensée à une partie de l'animal tels les hémisphères, que nous réduisons la personne à cette partie, et que nous soutenons que les parties d'animal sont de réelles entités, ayant une existence distincte de l'animal

lui-même. Cependant, si Parfit accepte l'affirmation de Johnston voulant que le prédicat *penseur* n'est utilisé correctement que lorsqu'il est attribué à un animal ou à une personne, comme il semble implicitement le faire (Parfit 2012 : 16), refuser la pensée à la personne reviendrait alors à l'accorder à l'animal. L'animal serait le penseur autonome plutôt qu'une de ses parties, et la personne un de ses sortaux de phase. Nous voilà donc de retour à l'animalisme classique.

En somme, il semble seulement cohérent d'attribuer la faculté de penser de façon autonome à une entité réelle, sans quoi il nous faudrait rejeter le *principe d'effectivité*. Cela signifie que si on est réductionniste à propos de la personne, il faut lui refuser cette faculté et toute possibilité de produire des effets dans le réel. Attribuer la pensée à la personne n'a de sens que dans le cadre d'un dualisme, puisque si la personne se réduit à une partie du corps tels les hémisphères, ce serait cette partie qui pense en propre plutôt que la personne. Par contre, il est plus intuitif d'avancer que c'est l'animal qui pense et que nous sommes numériquement identiques à des animaux, même s'il faut alors faire face au *problème des parties pensantes*. Toute partie de l'animal à laquelle on voudrait attribuer la pensée s'y trouverait d'ailleurs confrontée. Cependant, la thèse que c'est l'animal qui pense pourrait être supportée à l'aide d'un réductionnisme des parties d'animal, comme le défendent van Inwagen et Olson (voir 2.3.2), alors que cette voie ne serait pas disponible si l'on défend plutôt l'idée que c'est une partie de l'animal qui pense.

3.1.2 – Le problème des parties agissantes

Voyons maintenant une seconde objection majeure à la position finale de Parfit. Pour défendre la thèse que nous sommes la partie pensante de l'animal humain, Parfit ne voit pas de problème à retirer des facultés à l'animal pour les attribuer aux parties qui le constituent, qu'elles soient des parties corporelles tel l'estomac ou conceptuelles telle la personne. Après tout, l'animal possède encore ces facultés par l'entremise de ses parties ; il en hérite – ou ainsi se justifie Parfit. Le *principe du penseur non-dérivatif*, qui nous somme de nous identifier à l'entité qui pense par elle-même plutôt qu'à ce qui en hérite les facultés, soutiendrait d'ailleurs ce point de vue. Cependant, lorsqu'on examine le présupposé minimaliste derrière l'idée que c'est la partie qui pense plutôt que le tout, identifier cette partie se révèle ardu, au point où il n'est pas du tout clair qu'il n'y ait

qu'une seule partie pensante chez l'animal. Et lorsqu'on élargit le minimalisme aux facultés de l'animal autres que la pensée, nous sommes menés tout droit à un réductionnisme ontologique des animaux, et par suite de tout être vivant – une position grandement plus contestable que celle d'attribuer les facultés à l'animal plutôt qu'à ses parties.

Abordons le problème plus modestement, en nous en tenant à la faculté de penser pour le moment. Malgré que Parfit affirme que la personne est la partie pensante de l'animal, il semble qu'une version cohérente de sa position se réduise à l'affirmation que les hémisphères cérébraux sont cette partie, parce que ce sont eux qui réaliseraient nos facultés mentales. La question suivante se pose alors : pourquoi serait-ce les hémisphères qui sont retenus comme possesseurs de la faculté de penser ainsi que comme support de la personne, plutôt que le cerveau entier ou encore l'animal lui-même ? Pourquoi refuser à l'animal la faculté de penser de façon autonome et l'accorder exclusivement à cette partie du corps ? Le *principe du penseur non-dérivatif* justifie qu'on sélectionne cette partie comme le penseur que nous sommes à l'exclusion de l'animal et de toute autre partie *si l'on suppose que c'est cette partie qui pense par elle-même*. Mais pour affirmer que c'est bien la partie choisie qui pense de façon autonome, il faut une justification supplémentaire.

Olson suppose que le principe suivant est implicitement à l'œuvre, sans quoi le choix des hémisphères serait arbitraire :

Minimalisme du penseur (Thinking-Subject Minimalism) : « A true thinker has to be made up of all and only the objects directly involved in its thinking. » (Olson 2015a : 9)

Ce serait donc parce que l'animal, le cerveau entier et toute autre partie du corps comportent des éléments n'entrant pas directement dans la réalisation de la faculté de penser que cette faculté ne pourrait leur être attribuée. Olson pousse à l'absurde cette façon de régler le *problème des parties pensantes* :

« Having feet as parts is metaphysically incompatible with thinking. So my feet cannot be parts of me. » (Olson 2015a : 9)

Une telle position comporte de sérieux problèmes que nous allons présenter.

D'abord, souligne Olson, il n'est pas du tout clair qu'on puisse isoler ce qui est directement impliqué dans l'activité mentale d'un sujet. Notre activité mentale ne pourrait

avoir lieu sans irrigation sanguine qui alimente notre cerveau en oxygène et en nutriments, par exemple (Olson 2015a : 10). Notre cœur et nos poumons seraient-ils alors *directement* impliqués dans notre pensée ? Certes, leurs fonctions pourraient être remplies par des substituts artificiels, tel que le maintient Parfit, mais ces fonctions sont tout de même nécessaires. En limitant la partie pensante aux hémisphères cérébraux, Parfit semble dire que les fonctions circulatoires ne sont pas directement impliquées dans l'activité mentale du sujet ; mais son jugement ne va pas de soi. Sans critère précis pour distinguer ce qui est directement engagé dans une activité de ce qui ne l'est qu'indirectement, le minimalisme à propos du sujet pensant ne saurait être correctement applicable, ni être évalué quant à son bien-fondé. Attribuer la pensée exclusivement aux hémisphères serait donc prématuré.

Supposons tout de même que nous ayons un critère permettant de spécifier précisément ce qui est directement impliqué dans l'activité mentale d'un sujet, et que celui-ci écarte des éléments tels le cœur et les poumons. Aurait-on alors raison de croire qu'il n'y aurait qu'*une* partie pensante chez l'animal ? Ce n'est pas évident, puisque notre cerveau est composé d'aires spécialisées, chacune étant impliquée dans des fonctions différentes. Penser, pourrait-on dire, n'est pas une faculté mais plusieurs : la sensation visuelle engage un ensemble de neurones différent que la sensation auditive, même s'il peut y avoir des recoupements. Il en va de même pour la remémoration de visages et de noms, tout comme pour toute autre activité mentale (Olson 2015a : 11). Si l'on suit le *minimalisme du penseur*, selon lequel la partie pensante est uniquement constituée de ce qui entre directement dans la réalisation de l'activité mentale, alors chaque ensemble de neurones entrant dans *une* activité serait une partie distincte de l'animal, puisque même si ces ensembles se superposaient partiellement, ils ne seraient pas identiques. Olson conclut que ce principe nous mène à nier qu'il y a un penseur unique chez l'animal :

« It follows that a human being hasn't got one nonderivative 'general' thinker within it, but rather a crowd of narrow specialists: one that only imagines, another that only remembers, a third that does nothing but think about metaphysics, and so on. What we take to be a person able to perform all sorts of mental operations is really many beings, each able to perform only one. »
(Olson 2015a : 11)

Le *minimalisme du penseur* semble ainsi conduire à l'éclatement du penseur en une multitude de techniciens de la pensée, qui, d'ailleurs, ne se qualifieraient pas

individuellement comme des êtres conscients de soi et rationnels. Le minimalisme entraîne donc l'atomisation de la personne lockéenne – une conséquence qu'un lockéen comme Parfit ne pourrait accepter, préférant assurément refuser le principe. Cela ne correspond pas non plus à la phénoménologie de notre activité mentale, qui nous apparaît unifiée, ni à la conception de nous-mêmes qui en découle. Notre compréhension de nous-mêmes comme *penseurs* serait alors impossible.

Cela dit, peut-être qu'Olson est trop dur envers le minimalisme, et que la position de Parfit pourrait être justifiée sans fragmentation du penseur. Il est concevable qu'un critère minimaliste puisse faire concorder les parties de l'animal avec les divisions naturelles de notre corps en différents organes, muscles, et systèmes, par exemple, et rattacher les fonctions et activités à ces parties plutôt qu'à des ensembles toujours variables de neurones et d'autres cellules. Cela impliquerait peut-être que la partie pensante serait le cerveau plutôt que les hémisphères seuls. Mais quoique cela nous permettrait de sauver l'unité du *penseur*, cela ne saurait être une solution en ce qui concerne l'*agent*. En effet, lorsqu'afin d'être cohérent on généralise le *minimalisme du penseur* à toute activité de l'animal, les conséquences ne pourraient être simplement évitées en jouant sur la formulation du critère qui cerne les parties impliquées dans chacune de ses activités, comme nous allons le montrer.

Parfit est le premier à supposer que le *minimalisme du penseur* n'est qu'une application particulière d'un principe plus large. L'idée même que l'animal pense à l'aide d'une partie de son corps, il la base sur la supposition qu'il n'y a pas de controverse à considérer que l'animal remplit ses fonctions biologiques au moyen des parties de son corps, qui les accomplissent pour lui :

« Animals digest their food by having a part, their stomach, that does the digesting. Animals sneeze by having a part, their nose, that does the sneezing. These facts do not create a Too Many Digesters or Too Many Sneezer's Problem. Human animals think, we can similarly claim, by having a part that does the thinking. » (Parfit 2012 : 14-15)

Suivant Parfit, le minimalisme s'étendrait donc à toutes les activités de l'animal, la pensée n'en étant qu'une. Le principe général pourrait être formulé ainsi :

Minimalisme de l'agent : Le véritable agent d'une activité donnée est constitué de tous les objets directement impliqués dans cette activité, à l'exclusion de tout autre objet.

Autrement dit, l'agent d'une activité donnée serait la partie minimale du corps de l'animal qui l'exerce effectivement. L'animal serait donc divisé en une multitude d'agents minimaux étant chacun capable de n'accomplir qu'un seul type d'activité. L'animal serait ainsi constitué d'un agent marchant, d'un agent éternuant, d'un agent digérant, et ainsi de suite pour toutes ses activités.

Comme c'était le cas pour le penseur, il est possible de se demander si les agents que nous venons de nommer sont bien des agents *minimaux*, ou s'il faut les distinguer davantage. Est-ce le même agent minimal qui est impliqué dans la marche naturelle et la marche rapide, ou encore dans la digestion des sucres et des produits laitiers ? Les ensembles de cellules participant à ces activités seraient sûrement partiellement différents. Mais c'est là une distinction secondaire pour notre propos, et l'on peut continuer de supposer, pour être charitable, que les agents minimaux correspondent aux divisions naturelles du corps tels les organes. Ce qu'il importe de remarquer ici, c'est que le minimalisme implique que l'animal serait constitué d'un grand nombre d'agents minimaux, chacun correspondant à une partie distincte de l'animal, et ce peu importe le critère minimaliste utilisé pour cerner la partie directement impliquée dans l'exécution d'une activité donnée. Le *minimalisme de l'agent* a donc pour conséquence inévitable la fragmentation de l'*agent* que nous sommes en une multitude d'agents minimaux distincts.

Or, selon le *minimalisme de l'agent*, l'animal n'agit qu'en vertu de la partie – l'agent minimal – qui accomplit directement l'action pour lui. Par suite, chaque activité de l'animal serait en fait effectuée directement par un agent minimal, l'animal ne faisant qu'hériter de ses actions plutôt que de les réaliser en propre. Nous pourrions alors dire avec Parfit que lorsqu'un animal digère, c'est en fait son digérateur minimal, l'estomac, qui digère, alors que l'animal ne fait que digérer de façon dérivée – et pareillement pour toutes ses activités. Par conséquent, aucun animal n'agirait à strictement parler, ni ne posséderait de faculté d'agir, mais ne tiendrait ses actions et ses facultés que de ses parties proprement agissantes. Nommons cette conséquence le *problème des parties agissantes*.

Nous avons suggéré plus tôt que la position de Parfit lui permettait d'affirmer que l'animal pouvait *utiliser* ses parties pour agir, par exemple utiliser son cerveau pour penser, évitant alors une objection soulevée contre la position de Shoemaker (voir 2.3.3). Mais sa position telle qu'analysée ici s'avère incapable de faire même cela : *utiliser* ou

diriger des parties à agir, ce n'est là qu'un type d'activité pour lequel il y aurait un agent minimal correspondant, qui, *lui*, se chargerait directement de la coordination des actions. L'animal n'aurait donc même pas le contrôle de ses parties.

De plus, même si nous étions, comme Parfit le défend, la partie pensante et contrôlante de l'animal plutôt que l'animal lui-même, et que nous nous chargions de commander des actions aux autres agents minimaux, les actions qui en résultent ne seraient tout de même pas les nôtres, mais celles des parties qui les accomplissent. Les parties pensantes penseraient, point. À chaque partie appartiendrait son domaine d'influence propre. Si mes pieds marchent, et que mes cordes vocales produisent des sons, alors moi, la personne incarnée, ne pourrait marcher ni parler. Lorsqu'on mène la *Embodied Person View* à sa conclusion, la personne en vient jusqu'à perdre la propriété de son corps et son effectivité physique. Pourrait-on alors affirmer de façon cohérente qu'elle serait responsable des gestes que les autres parties logeant dans le même animal posent ? Les comportements n'étant imputables qu'aux membres du corps qui les mettent en mouvement, toute justice serait totalement absurde.

Bref, les conséquences du *minimalisme de l'agent* sont désastreuses pour notre compréhension du réel. L'animal, plutôt que de se mouvoir de lui-même, se trouverait réduit à un automate qui est agité par ses parties. Le minimalisme élimine ainsi la possibilité d'une compréhension de l'animal comme *agent*, son agentivité s'avérant épiphénoménale. Plus encore, s'il implique que toute entité n'agit que par l'entremise d'un essaim d'agents minimaux hautement spécialisés, alors *nous* – que nous soyons identiques à des animaux, des personnes, ou une partie d'animal donnée – ne pourrions plus être compris comme des agents unifiés. Et puisque qu'il ne semble pas y avoir de raison qui justifierait qu'on applique le minimalisme aux animaux mais pas aux autres entités constituées de parties, aucun être vivant – ou plutôt aucun être *agissant*, vivant ou non – ne serait *un agent*. Expliquer des ensembles d'actions concertées en termes de motivations et d'intentions de l'agent serait alors ardu voire impossible. Sans agent unifié pour ancrer l'activité de chaque être vivant, le réel serait alors un lieu où se jouent des actions sans sens ni dessein.

Finalement, non seulement la position finale de Parfit nous retire-t-elle la possibilité de nous comprendre comme *agents* unifiés, et peut-être même comme

penseurs unifiés, elle a également d'inquiétantes implications ontologiques. En effet, si tout animal et par extension tout être vivant est absolument impotent, puisqu'il ne peut réaliser d'activité par lui-même, alors qu'est-ce qui nous permet de dire qu'il aurait une réelle existence ? Reste-t-il quelque chose de l'animal lorsqu'on le distingue de toutes les parties de son corps ainsi que des facultés qui leurs sont associées ?

Quoiqu'il ne soit pas évident de fournir un critère qui distingue ce qui est réel de ce qui ne l'est pas, la réciproque du *principe d'effectivité* a plusieurs adhérents, dont Jaegwon Kim (1998 : 119) :

Principe de réalité : Seules des entités ayant des pouvoirs causaux et pouvant produire des effets réels ont une existence réelle.

Autrement dit, être réel, c'est pouvoir faire une *différence causale* par sa présence ou son absence (Kim 1998 : 31). Parfit souscrit également à ce principe, quoiqu'il ne le formule pas ainsi. Rappelons que, selon lui, une entité dont l'existence consiste en un *further fact* que l'existence de ses constituants a une existence réelle, puisque son activité ferait une différence supplémentaire quant à ce qui se passe dans le réel (Parfit 2007 : 5-6). Si ce n'est pas le cas, alors l'existence de l'entité se réduit à celle de ses constituants, et les faits à son propos se réduisent aux faits à propos de ces derniers. Des entités et des faits ainsi réduits parce qu'ils ne font aucune différence réelle sont alors selon Parfit de nature conceptuelle ; c'est là le cœur de son réductionnisme constitutif.

Suivant un tel *principe de réalité*, l'animal ne fait aucune différence causale par lui-même, mais seulement par l'entremise des parties qui le constituent. N'ayant aucune effectivité propre, il n'aurait donc aucune réalité et ne serait alors qu'une entité conceptuelle, n'existant que parce que le concept entre dans notre usage courant. Son existence se réduirait alors à celle des parties qui le constituent – la pléthore d'agents minimaux – qui, *elles*, seraient des entités ayant une réelle existence, puisqu'elles seules sont capables de poser des actions ayant une efficacité causale. Parfit n'affirme jamais explicitement une telle conception des animaux, quoiqu'il la suggère, sans toutefois la mener à sa conclusion :

« The thinking part of a human animal, we could say, is related to this animal's cerebrum or upper brain in a way that is roughly similar to that in which this animal is related to its whole body. » (Parfit 2012 : 15)

L'existence de la partie pensante de l'animal, c'est-à-dire de la personne, se réduit selon

Parfit à celle de ses constituants, les hémisphères cérébraux dans ce cas – ce qui en fait une entité conceptuelle, tel que Parfit le maintient durant toute sa carrière. Mais si la relation de l'animal à son corps est analogue à celle de la personne à ses hémisphères, cela signifie que l'existence de l'animal se réduit pareillement à celle de son corps, c'est-à-dire, pouvons-nous ajouter, à l'ensemble des parties du corps qui le constituent. L'animal ne serait donc qu'un *bundle* conceptuel de parties corporelles. Le réductionnisme constitutif de Parfit s'appliquerait donc tout autant à la personne qu'à l'animal.

Par conséquent, la position finale de Parfit implique que les animaux n'ont pas d'existence réelle – et pareillement pour tout être vivant. Ils ne seraient pas des objets physiquement localisés dans le monde, mais seulement dans notre schème conceptuel. Dans notre ontologie, il n'y aurait ainsi pas d'animaux ni de personnes, mais seulement des parties d'animaux juxtaposées, c'est-à-dire des nuées d'agents minimaux. Ce serait un monde bien bizarre à habiter.

À la lumière d'une telle position, il semble bien plus sobre d'avancer que nous sommes numériquement identiques à des animaux, que les animaux et les autres êtres vivants existent réellement, qu'ils pensent et agissent par eux-mêmes, et de se charger des *problèmes des parties pensantes et agissantes* en défendant l'idée que les parties d'animaux n'ont pas d'existence indépendante, mais entrent dans la composition du tout qu'est l'animal.

*

En somme, la *Embodied Person View* de Parfit fait face à de sérieuses objections. D'abord, il semble bien difficile de justifier l'attribution de la faculté de penser de façon autonome à une entité qu'on a ontologiquement réduite à d'autres. Si la personne, l'animal, ou les parties de son corps ne sont pas des entités réelles, alors ils devraient être éliminés de la liste de candidats au titre de penseur. Comme nous l'avons avancé, la personne est le candidat le moins plausible de la liste, requérant soit l'adoption d'une forme de dualisme, soit la réduction d'une partie physique de l'animal à la personne – une stratégie discutable.

Ensuite, si l'on tente d'attribuer la pensée à une partie du corps de l'animal plutôt qu'au tout, il faut alors pouvoir justifier la délimitation d'une telle partie. Le *minimalisme de l'agent* semble être le principe derrière l'identification des hémisphères cérébraux

comme la partie proprement pensante. Cependant, il est particulièrement difficile de déterminer ce qui entre *directement* dans la réalisation d'une activité donnée de l'animal. Mais peu importe le critère utilisé, le minimalisme conduit au morcellement de l'*agent* que nous sommes, voire même du *penseur*. Et si toute action était en fait directement réalisée par un agent minimal qui n'est pas l'animal, les animaux seraient privés de toute effectivité, et réduits à une existence purement conceptuelle. C'est là le *problème des parties agissantes*.

Les animaux ne seraient ainsi conçus que comme des agglomérats de parties et de propriétés qui ne sont pas à strictement parler les leurs. Non seulement ne penseraient-ils et n'agiraient-ils que de façon dérivée, leur existence serait elle-même parasitaire sur celle de leurs parties. Ils ne seraient donc pas des objets matériels. Si la position de Parfit visait à expliquer le lien entre nous, la personne à son avis, et l'animal, et ce en réponse aux critiques animalistes et dans le but de nous rattacher à un objet matériel, le moins qu'on puisse dire c'est que le succès est loin d'être atteint. Il serait certes possible d'émettre d'autres objections à la position finale de Parfit, mais celles que nous avons présentées ici lui posent déjà suffisamment de difficultés. Tout bien considéré, l'animalisme semble demeurer la position la plus plausible.

3.2 – Un argument pour un animalisme compatibiliste

Dans les chapitres précédents, nous avons présenté les positions lockéennes de Parfit et de Shoemaker ainsi que la position animaliste classique que défendent Olson et Snowdon. Nous avons examiné les problèmes de chacune. Le lockéanisme sous ses diverses formes nous conduisait à affirmer que nous n'avons jamais été des fœtus ; le constitutionnalisme que Parfit adoptait initialement impliquait qu'il y a deux sujets pensants coexistant dans chaque être humain ; les positions initiale et finale de Parfit faisaient de nous des penseurs irréels et réifiaient les parties du corps animal ; et l'animalisme classique rejetait la *Transplant Intuition* mais avait pour conséquence qu'une personne-résiduelle serait créée suite à la séparation du cerveau d'avec le reste du corps. La seconde position de Parfit, qui faisait converger les critères d'identité animale et personnelle en ce qu'ils nécessitaient tous deux la continuité du cerveau, nous semblait

prometteuse, mais puisque Parfit rejetait ses implications ontologiques, elle n'était ni achevée ni construite de façon cohérente.

Dans ce chapitre, nous allons dresser le portrait d'une position qui mérite à notre avis d'être considérée plus avant. Pour ce faire, nous allons prendre comme cadre et comme contraintes les arguments que nous avons vus jusqu'ici, et tenter de formuler une position plausible qui pourrait éviter les problèmes principaux des positions que nous avons examinées jusqu'ici – quoiqu'elle ait sans doute les siens. Nous allons chercher ce faisant à réconcilier l'animalisme et le lockéanisme en une position cohérente qui conserve maints de leurs apports. La position résultante se rapproche donc de la seconde position de Parfit, quoique c'en est une version animaliste plutôt que lockéenne. On peut ainsi dire que ce serait un *animalisme compatibiliste*. Nous allons procéder en présentant d'abord la position (3.2.1), pour ensuite fournir quelques arguments en appui (3.2.2-3.2.4), insuffisants peut-être, mais qui pourraient servir de point de départ pour une défense ultérieure.

3.2.1 – L'animalisme compatibiliste

Commençons par disposer devant nous les pièces de notre casse-tête afin de se faire une image de la position compatibiliste qu'elles peuvent former une fois assemblées.

L'argument du fœtus (2.1.1) élaboré par les animalistes montre que si nous croyons que nos conditions de persistance sont psychologiques, nous ne pouvons avoir été des fœtus parce que ceux-ci ne commencent pas leur vie avec des propriétés psychologiques. L'approche psychologique de l'identité implique que la personne lockéenne est une entité numériquement distincte de l'animal humain, qui se joint à lui ou surgit en lui plus tard durant son développement. Si nous voulons affirmer que nous avons été des fœtus plutôt qu'un parasite lockéen qui s'est accaparé de notre cerveau, il nous semble devoir alors admettre que nos conditions de persistance sont biologiques, et que nous sommes par suite numériquement identiques à l'animal humain que nous voyons dans le miroir.

Le constitutionnalisme se propose comme solution à ce problème en suggérant qu'il est plausible de considérer que deux êtres numériquement distincts partageant la même matière si l'un entre dans la constitution de l'autre. Cependant, cette position fait

face à un nouveau problème : celui de la multiplication des penseurs, personnes, et êtres vivants. *L'argument de l'animal pensant* (2.1.2) expose ce problème et nous conduit à n'accorder la pensée et le statut de personne qu'à une seule entité d'un même type dans un même espace. Cela découle du *principe d'exclusion* (2.1.2), selon lequel s'il était concevable que plusieurs entités du même type occupent le même espace et soient constituées de la même matière, et qu'elles aient ainsi toutes leurs propriétés en commun, il serait alors impossible de les distinguer les unes des autres. Cette contrainte nous somme donc de n'attribuer la pensée et le statut de personne qu'à une seule entité composant l'être humain, fût-elle l'animal lui-même, une partie de l'animal, ou la personne.

Or, nous avons défendu l'idée que pour pouvoir exécuter des activités telle la pensée, il faut que l'entité ait une existence réelle, plutôt que d'être ontologiquement réductible à d'autres entités et de n'avoir alors qu'une existence conceptuelle (3.1.1). Une entité ainsi réduite ne penserait pas par elle-même, mais seulement en vertu d'une ou des entités à laquelle elle est réduite. Dans un tel cas, il serait plus plausible d'attribuer la pensée à ces dernières. C'est là ce que prescrit le *principe d'effectivité*.

Quelles seraient alors les entités réelles peuplant notre monde ? Autrement dit, à quelles entités le réductionnisme ontologique ne s'appliquerait-il pas ? Cette question est plus difficile. Sa portée est large et dépasse le cadre de ce mémoire, malgré qu'une réponse semble nécessaire pour déterminer ce que nous sommes. Nous avons suggéré que la position de van Inwagen – selon laquelle, outre les particules simples, seuls les êtres vivants seraient des objets réels complexes – mérite d'être considérée (1.2.2). Si le réductionnisme s'appliquait universellement, tous les objets seraient conceptuels, et nous nous verrions forcés d'attribuer la pensée à des ensembles de particules. Par contre, si le réductionnisme ne s'appliquait jamais, et que la composition de particules en objets réels avait toujours lieu, alors les mêmes particules composeraient une multitude d'objets réels numériquement distincts. Les particules composant nos hémisphères cérébraux composeraient également notre cerveau et l'animal entier. Ce serait problématique. Il semble néanmoins plausible de croire que si le réductionnisme universel était faux, les êtres vivants seraient les entités les plus susceptibles d'avoir le statut d'entité réelle.

Cependant, même sans théorie de la composition pour nous aider dans

l'attribution de la pensée à une entité réelle, d'autres arguments peuvent servir à départager les candidats. Si nos propositions jusqu'ici ne sont pas erronées, alors à moins d'être dualiste, la personne ne serait pas admissible si on la conçoit comme le fait Parfit, c'est-à-dire comme une entité conceptuelle. Il nous semble donc devoir déterminer, entre l'animal et une infinité de subdivisions possibles de l'animal, quel serait le penseur autonome. Ceux-ci étant constitués ou composés de la même matière qui réalise la pensée, ils semblent tout autant éligibles les uns que les autres ; c'est ce qu'Olson nomme le *problème des parties pensantes* (2.3.2), que l'animalisme a bien de la difficulté à dénouer. Olson avance que l'ontologie de van Inwagen, qui est réductionniste à propos des parties de l'animal, serait la meilleure pour écarter tous les candidats autres que l'animal. Nous avons remarqué que même si cette position pouvait paraître draconienne, quelqu'un qui défend que c'est plutôt une partie de l'animal qui pense se trouverait confronté au même problème, devant écarter à la fois l'animal et toutes les autres parties de l'animal, sans toutefois avoir accès à la solution van-inwagienne, puisqu'elle réduirait également la partie à laquelle il voudrait attribuer la pensée (3.1.1). De plus, nous avons défendu l'idée que si nous attribuions aux parties de l'animal les activités qu'on reconnaît normalement à l'animal, cela reviendrait à ôter toute effectivité à l'animal, autrement dit à être réductionniste à propos des animaux et des êtres vivants de façon générale (3.1.2). C'est ce que nous avons nommé le *problème des parties agissantes*.

En somme, il semble que si nous considérons l'animal réel et voulons lui accorder la pensée, il nous faut refuser la réalité de ses parties, alors que si nous considérons plutôt que ce sont les parties qui sont réelles et qui accomplissent des activités telle la pensée, il nous faut refuser la réalité de l'animal. La réalité de l'animal et de ses parties seraient apparemment mutuellement exclusives. Devant un tel dilemme, il semble plus intuitif d'accorder la réalité et la pensée à l'animal. D'ailleurs, si on voulait l'accorder à une partie, il faudrait fournir un argument justifiant la délimitation de celle-ci et la réduction des autres parties tout autant éligibles. Qu'est-ce qui expliquerait qu'une subdivision donnée de l'animal soit une entité réelle, mais que cette même subdivision à laquelle on ajouterait une cellule avoisinante n'en serait pas une ? Le *minimalisme du penseur* ou de *l'agent* serait alors une avenue à explorer, mais, comme Olson l'a indiqué, bien des organes et des systèmes entrent dans la réalisation d'une activité donnée, rendant la

possibilité d'une délimitation claire du penseur ou de l'agent minimal douteuse (3.1.2). Si on amène cette critique à sa conclusion ultime, il pourrait très bien s'avérer que le penseur et l'agent minimal soit l'animal lui-même, ce dernier étant un tout fonctionnellement interdépendant.

Compte tenu de tous les arguments avancés, il paraît plus plausible de donner gain de cause à l'animal, et de considérer que ce soit lui qui a une existence réelle et qui détient en propre la faculté de penser, plutôt qu'une de ses parties tel le cerveau. Or, suivant le *principe du penseur non-dérivatif* (2.3), que Parfit et Olson adoptent tout deux, nous serions numériquement identiques au penseur autonome plutôt qu'à une entité qui en hériterait les facultés de façon dérivée. Nous serions donc l'animal humain. Le concept d'*animal* ou d'*animal humain* ferait donc office de sortal de substance pour nous, déterminant que nous avons des conditions de persistance biologiques, tel que l'affirment les animalistes (2.1.3). Cela s'aligne avec l'idée de départ que nous avons commencé notre existence comme fœtus, et que ceux-ci ne peuvent avoir de conditions de persistance psychologiques.

Cela dit, si nous sommes l'animal humain, qu'en est-il alors de la personne ? Puisqu'on a convenu que c'est l'animal qui est réel et qui pense en propre, c'est lui qui peut avoir les propriétés de conscience de soi et de rationalité que Locke retient comme essentielles à la personne. Ce serait donc l'animal qui est une personne. Si on affirmait que la personne était plutôt une entité numériquement distincte de lui, ayant les mêmes propriétés mentales qui en font une personne, on serait de nouveau pris avec les problèmes du constitutionnalisme, notamment celui de la prolifération des penseurs et des personnes. Il faut donc être pleinement réductionniste à propos des personnes, et concéder qu'elles tiennent leurs propriétés mentales de l'entité réelle qui les a en propre, l'animal humain dans notre cas.

Or, comme l'animal humain n'a pas de propriétés mentales durant toute sa vie, la débutant comme fœtus et pouvant tomber de façon permanente dans un état végétatif, il semble approprié de considérer que, comme l'avancent les animalistes, la personne serait un *sortal de phase*, c'est-à-dire un qualificatif attribuable à l'animal humain durant une phase de sa vie, et potentiellement à d'autres entités ainsi psychologiquement douées (2.1.3). La personne ne serait donc pas une sorte d'entité fondamentale au même titre que

les espèces animales. De plus, puisque son existence et ses propriétés se réduiraient à celles de l'animal, il est cohérent de considérer qu'elle n'existe que tant que l'animal dont elle dépend continue d'exister. Quoique ça puisse sembler vite dit, cela implique que la carrière de la personne ne pourrait diverger de celle de l'animal, contrairement à ce qu'en croient les lockéens et même des animalistes comme Olson, et à ce que prétend montrer la multitude d'expériences de pensée qu'ils mobilisent. Il semble contradictoire de réduire l'existence d'une entité tout en maintenant qu'elle pourrait survivre sans sa base de réduction. La personne vivrait et mourrait donc avec l'entité réelle à laquelle elle se rattache. Nous reviendrons sur cette affirmation controversée plus loin.

Par contre, quoique la personne ne puisse être numériquement distincte de l'animal, nous ne voyons pas de mal à ce qu'on la conçoive comme une *entité* qui en serait *conceptuellement* distincte plutôt qu'un simple prédicat, comme Parfit continuait de le faire lors de son rapprochement avec l'animalisme (2.2.1). Cela permettrait de traiter la personne séparément et de lui assigner son propre critère d'identité – relative certes, puisqu'à propos d'une entité sous une description particulière – qui consisterait en la continuité de l'existence de l'animal *et* de ses propriétés mentales qui en font une personne. Si toutefois un argument montrait que cette manière de traiter la personne pose problème, on pourrait très bien y renoncer, celle-ci n'étant qu'un aspect marginal de la position compatibiliste.

Il semblerait que le portait de l'animalisme compatibiliste dressé jusqu'ici ne soit qu'une reprise de l'animalisme classique que défend Olson, mis à part quelques détails. Il y a effectivement bien des similarités, et c'est d'ailleurs pour cette raison que c'est avant tout un animalisme. Cependant, malgré que la personne ne soit pour Olson qu'un sortal de phase et que ses facultés appartiennent en fait à l'animal, sa position implique tout de même que la carrière de la personne pourrait diverger de celle de l'animal, si on sépare le cerveau (ou les hémisphères cérébraux) du corps et qu'on fournit à chacun un système de soutien de la vie. *Nous* irions là où va notre corps décérébré, et la personne, dont l'existence était réduite à celle de l'animal, acquerrait alors une réelle existence, numériquement distincte de celle de l'animal, et irait là où va le cerveau. C'est le *problème de la personne-résiduelle* (2.2.2). L'animalisme d'Olson apparaît ainsi comme l'envers du lockéanisme, tous deux considérant qu'animal et personne seraient séparables.

La différence entre eux consisterait en ce que de tels animalistes affirment que nous irions avec le corps décérébré, alors que pour les lockéens nous irions plutôt avec notre cerveau ou nos hémisphères. Ce n'est pas une conséquence acceptable si on réduit pleinement la personne à l'animal et qu'on veut que notre position soit cohérente.

Apportons quelques précisions à ce sujet. Olson affirmait originellement que l'organisme irait là où va le cerveau entier, mais pas là où iraient les hémisphères cérébraux seuls, parce qu'il croyait que le tronc cérébral était essentiel au maintien des fonctions auto-régulatrices de l'organisme (2.2.3). Autrement dit, il défendait la *Whole Brain Transplant Intuition*, mais rejetait la *Cerebrum Transplant Intuition*. Cependant, après avoir pris connaissance des arguments de Shewmon, selon lequel la mort cérébrale n'implique pas nécessairement la mort de l'organisme parce que le cerveau ne joue pas le rôle d'intégrer ses fonctions auto-régulatrices (2.2.3), Olson rejette également la *Whole Brain Transplant Intuition*. Selon lui, si Shewmon a raison, alors nous, l'animal humain, resterions avec le corps décérébré suite à une transplantation de notre cerveau entier, parce que le cerveau n'est qu'un organe et pas un organisme (Olson 2016 : 6-7). Il y a certes la possibilité d'un cas de fission, où le cerveau entier *et* le corps décérébré survivant séparément seraient tous deux considérés comme des organismes animaux, qui pourrait remettre en question sa position. Olson juge cependant qu'un tel cas repose sur des assomptions douteuses, à savoir que chacune des parties serait un organisme, et que chacune serait l'organisme original si l'autre partie était absente (Olson 2016 : 8).

Ainsi, en acceptant la séparabilité de la personne d'avec l'animal, la position d'Olson va non seulement à l'encontre de l'idée que l'existence de la personne se réduirait à celle de l'animal et qu'elle n'en serait pas numériquement distincte ; mais elle traite également la conscience phénoménale et les facultés mentales comme non-essentiels au maintien de notre identité animale. Olson se justifie à l'aide de l'*argument du fœtus*, soutenant que puisque ces facultés ne sont pas présentes au début de notre existence, elles ne peuvent faire partie de notre critère d'identité animale (2.1.1). De plus, intégrer ces facultés à notre critère d'identité ferait en sorte que, selon lui, nous n'aurions pas nos conditions de persistance en commun avec tous les animaux ; nous ne les tiendrions pas en vertu du fait que notre sort de substance est l'*animal*, mais au nom de notre appartenance à une espèce animale particulière dotée de facultés propres (2.1.3). Nous

reviendrons sur ce problème plus loin (3.2.2).

Afin d'éviter de tels problèmes et incohérences dans la position que nous esquissons ici, il nous faut prendre au sérieux l'inséparabilité de la personne d'avec l'animal qu'implique le réductionnisme ontologique de la personne, et voir où cela nous mène. Il n'y aurait alors qu'une seule possibilité acceptable : lorsqu'on divise l'animal humain, la personne irait avec la partie qui préserve la vie de l'animal, si cette partie demeure également capable de préserver l'identité personnelle. Il ne pourrait donc y avoir de *personne-résiduelle*, c'est-à-dire une personne humaine qui n'est pas également un animal humain. Par suite, si, comme l'évoquent les expériences de pensée, le cerveau pouvait survivre une fois détaché du corps et continuer de supporter la conscience moyennant un système de soutien de la vie, alors l'animalisme compatibiliste impliquerait que l'animal et la personne iraient avec le cerveau. Notons que cela ne signifie pas que l'animal et la personne seraient en fait une partie de l'animal ; au contraire, pour affirmer que l'animal est l'entité réelle et pensante, l'animalisme compatibiliste semble devoir maintenir que les parties de l'animal n'ont pas d'existence propre, comme nous l'avons souligné plus haut. Par suite, le cerveau détaché – ou plus précisément suffisamment du cerveau pour supporter la conscience – serait en fait l'organisme animal maximale ment mutilé, plutôt qu'un simple organe comme le défend Olson.

Pour être plausible, l'animalisme compatibiliste devrait pouvoir justifier cette affirmation. Une telle justification ne devrait pas faire appel à des faits psychologiques, sans quoi notre critère d'identité ne pourrait être biologique. Une piste possible serait de défendre que le cerveau est le support de facultés essentielles à l'identité des animaux humains. Nous avons suggéré plus tôt, en présentant la seconde position de Parfit, que l'intentionnalité pourrait être une de ces facultés (2.2.3). La conscience phénoménale pourrait en être une autre. En quoi ces facultés seraient biologiques, et en quoi elles seraient essentielles, nous en discuterons dans les sous-sections suivantes (3.2.2-3.2.3).

Toutefois, si l'animalisme compatibiliste souscrit à la *Whole Brain Transplant Intuition*, c'est une autre histoire en ce qui concerne la *Cerebrum Transplant Intuition*. En effet, suivant les recherches de Pinto et son équipe que nous avons présentées dans le premier chapitre (1.3.2), il y a des raisons de croire que les hémisphères cérébraux ne pourraient maintenir la conscience à eux seuls. Pinto a montré que lorsque les

hémisphères sont séparés, c'est-à-dire que le corps calleux les reliant est coupé, la conscience demeure unifiée. Un hémisphère sait ce que l'autre voit sans même avoir accès à l'information visuelle que véhicule le corps calleux, indiquant que l'unité de la conscience est assurée par autre chose que les hémisphères pris individuellement ou le corps calleux qui les relie. Il serait donc tout à fait possible qu'une autre partie du cerveau, le tronc cérébral par exemple, soit essentiel pour maintenir la conscience. À moins de preuve du contraire, il semble judicieux de rejeter la *Cerebrum Transplant Intuition*.

Cela dit, il nous reste un dernier problème avec lequel l'animalisme compatibiliste doit composer : la possibilité de la fission de l'organisme. Si Shewmon avait raison, et qu'aucun organe n'était essentiel au maintien de la vie de l'organisme, il serait alors concevable que lorsqu'on sépare le cerveau du reste du corps, on ait un cas de double survie, chaque partie étant considérée comme un organisme à part entière. Si l'animalisme compatibiliste identifie l'animal original et la personne au cerveau, cela signifie qu'un deuxième animal serait créé lorsque le cerveau est séparé du reste du corps. C'est le problème de l'*animal-résiduel* (2.2.3).

D'abord, il pourrait bien se trouver que Shewmon ait tort. Malgré que ses recherches aient été prises au sérieux, le consensus demeure de l'avis que la mort cérébrale est la mort de l'organisme, et ce même si on ne considère plus que le cerveau intègre les fonctions auto-régulatrices de l'organisme (President's Council on Bioethics 2008 : 89-91). On pourrait alors suivre van Inwagen et considérer le corps décérébré comme du simple tissu vivant plutôt que comme un organisme, même s'il continue de maintenir certaines fonctions biologiques. Perdre notre corps à l'exception de notre cerveau serait analogue à perdre un doigt, mis à part les traumatismes encourus et le système de soutien de la vie nécessaire pour retrouver une fonctionnalité normale. Justifier que notre cerveau assure la continuité de facultés essentielles à l'organisme que le reste du corps ne peut préserver serait un appui nécessaire à cette position.

Par contre, si considérer le corps décérébré comme du tissu vivant n'était pas une position admissible, et que ce corps était effectivement un organisme, cela poserait problème pour l'animalisme compatibiliste. Il serait certes possible de défendre l'idée que sans les facultés essentielles que préserve le cerveau, cet animal-résiduel ne serait pas

l'organisme original, parce qu'il ne serait plus un animal *humain*, voire qu'il ne serait plus un *animal* tout court. Par contre, cela impliquerait que tout être humain serait en fait deux organismes étroitement liés – une conséquence peu plausible. Ce problème n'est toutefois pas réservé au compatibilisme mais affecte toute position rivale. Si le corps décérébré était un organisme, toute position devrait admettre que chaque être humain serait en fait deux entités qui coexistent : soit deux organismes, soit un organisme et une personne, selon le sortal de substance attribué au cerveau détaché. Il n'y a peut-être pas d'alternative au minimalisme ontologique de van Inwagen pour éviter le problème de l'animal-résiduel.

Voilà qui nous donne une bonne idée de ce que serait un animalisme compatibiliste. En bref, nous serions une entité ayant une existence réelle, l'animal humain, qui peut penser et être une personne durant une partie de sa vie. Nos conditions de persistance seraient biologiques, et requerraient la continuité de suffisamment de notre cerveau pour supporter la conscience, qui serait essentielle à notre identité humaine. Nous pourrions ainsi être réductionnistes à propos des personnes tout en étant non-réductionnistes à propos de ce que nous sommes.

Nous allons à présent donner quelques appuis à cette position – sur lesquels il faudrait élaborer lors de recherches ultérieures – ainsi que quelques critiques. Nous allons d'abord discuter de l'importance du cerveau et de la conscience phénoménale (3.2.2) ; ensuite, de la manière de rendre leur présumée nécessité pour notre identité compatible avec le fait que nous avons été des fœtus (3.2.3) ; et finalement, de la possibilité de séparer le contenu mental de la conscience qui la rend possible (3.2.4).

3.2.2 – Le cerveau et la conscience phénoménale

L'animalisme compatibiliste dont nous venons de donner un aperçu implique que la partie essentielle de l'organisme humain serait son cerveau. Afin que cette position soit cohérente et recevable, un compatibiliste devrait pouvoir justifier les affirmations suivantes : 1) que la continuité de l'organisme humain requiert plus que la continuité de ses fonctions auto-régulatrices ; 2) qu'il y a des fonctions essentielles à l'organisme et à sa survie qui sont assurées par le cerveau ; et 3) que ces fonctions peuvent être intégrées à un critère d'identité biologique. C'est seulement si ces trois propositions sont

convenablement appuyées qu'il serait plausible de croire que le cerveau pourrait être la partie essentielle de l'organisme humain, et de considérer l'animalisme compatibiliste comme une position viable. Nous ne pouvons fournir une telle justification ici, et d'ailleurs les données empiriques dont nous disposons actuellement peuvent paraître insuffisantes pour le faire de façon robuste. Nous allons nous contenter de formuler quelques arguments allant en ce sens. Dans cette section, nous allons nous concentrer sur la deuxième proposition ; la troisième sera l'objet de la suivante.

Nous avons avancé plus tôt qu'il y a des raisons de croire que le cerveau assure des fonctions essentielles à l'organisme. Notamment, nous avons proposé l'idée que l'intentionnalité serait une faculté vitale pour les organismes motiles, parce qu'ils doivent diriger leur activité sur leur environnement extérieur pour assurer leur subsistance (2.2.3). Nous avons suggéré que chez des animaux dotés d'un cerveau, il est plausible que cette faculté soit rattachée à cet organe. Nous allons explorer ici une autre avenue que nous jugeons pertinente : l'idée que la conscience phénoménale, qui semble tout autant rattachable au cerveau, ferait partie intégrante de l'identité des organismes dotés de cette faculté. La formulation que Nagel donne de cette faculté est utile ici :

« [T]he fact that an organism has conscious experience *at all* means, basically, that there is something it is like to *be* that organism [...] – something it is like *for* the organism. » (Nagel 1974 : 166)

Autrement dit, la conscience phénoménale est ce à quoi on réfère lorsqu'on parle d'organismes conscients, et est ce qui en fait des *sujets d'expérience*, capables d'avoir des expériences qualitatives tels le plaisir et la douleur. Tentons de faire un pas vers l'idée qu'elle serait essentielle aux organismes ainsi pourvus.

Débutons la discussion en nous aventurant sur le terrain de jeu des auteurs : celui des expériences de pensée. Nous discuterons des données scientifiques par la suite. Voyons quelles sont nos intuitions lorsqu'on imagine une transplantation de cerveau chez des animaux non-humains qu'on considère couramment conscients. Prenons le cas suivant. Votre vétérinaire vous annonce que Rex, votre chien et fidèle compagnon depuis déjà dix ans, est atteint d'une maladie mortelle et inguérissable. Il vous propose le seul traitement disponible : greffer son cerveau encore sain à un clone, ce qui vous permettrait de poursuivre votre relation avec lui. Si l'opération était réussie, le vétérinaire aurait-il seulement déplacé un organe, ou sauvé l'animal en lui donnant un nouveau corps ?

En supposant que la vie mentale de Rex serait préservée, la *Transplant Intuition* ne perd pas de sa force : on s'attendrait à ce que l'animal résultant vous reconnaisse, qu'il connaisse les habitudes et les jeux que vous avez cultivés avec Rex, et que son comportement général soit en continuité avec celui que Rex avait avant l'opération ; bref, on s'attendrait à ce qu'il *soit* Rex. Toutefois, les positions de Parfit et d'Olson impliquent toutes deux que ce ne serait pas le même animal. Parfit avance que ce ne serait que sa partie pensante, numériquement distincte de l'animal original, qui serait transférée :

« [W]e can also apply the Embodied Part View to such animals as dogs, whose conscious, thinking and controlling part may be at best a borderline case of a Lockean person. » (Parfit 2012 : 17, note 20)

Harold Noonan remarque que l'animalisme classique d'Olson et Snowdon implique que ce ne serait qu'un organe qui est préservé dans ce cas, plutôt que l'organisme :

« [S]ince dogs also begin life as fetuses and may end it in a permanent vegetative state, these considerations provide equally strong support, *mutatis mutandis*, for distinguishing psychologically individuated canine individuals ('cursons') from canine animals. » (Noonan 2001 : 86)

De telles positions, selon lesquelles les animaux conscients seraient en fait deux entités séparables, l'animal et ce qu'on peut considérer être le *sujet d'expérience* – puisqu'il serait incorrect d'accorder le plein statut de personne aux animaux non-humains – ont un vaste impact sur notre conception de ces êtres vivants.

D'abord, si Parfit et Olson avaient raison, l'être avec lequel vous avez entretenu une longue amitié et qui répond au nom de « Rex » ne serait pas en fait un chien, mais bien son organe pensant, le sujet d'expérience. Quoique son comportement soit typiquement celui d'un chien, il ne serait un chien que de façon contingente. Si on généralise, ceux qu'on croit être des animaux conscients ne seraient en fait que des tranches incarnées d'animaux, des organes pensants sur pattes. Les vrais animaux, quant à eux, ne seraient que de complexes machines biologiques, l'équivalent d'un système de soutien de la vie sophistiqué pour organes conscients. Ils ne disposeraient d'aucune autonomie sans être liés à une entité qui en prendrait le contrôle. C'est toute une chance que la nature ait si bien fait les choses que de tels animaux naissent systématiquement assemblés à une partie consciente qui les complète.

Une telle conception des animaux qu'on juge conscients serait très limitée, c'est le moins qu'on puisse dire. Les positions de Parfit et d'Olson impliquent non seulement

qu'ils n'ont pas de vie psychologique, mais que ni la sensibilité ni l'émotivité ne font partie intégrante de ces animaux ; ces facultés ne leur seraient pas essentielles et n'entreraient pas dans leur identité. Ainsi, cette conception des animaux serait foncièrement mécaniste, l'activité de l'animal se trouvant réduite à un ensemble de réflexes et d'automatismes inconscients. Autant pour Olson que pour Parfit, cette mécanisation de l'animal découle de l'idée que ce qui peut avoir des états conscients serait une partie dissociable de l'animal et donc numériquement distincte de lui. Pour Olson, c'est le cas parce que le critère d'identité biologique des animaux n'inclut ni le cerveau ni la conscience qui s'y rattache ; alors que pour Parfit, c'est parce que la partie consciente appartient exclusivement à la personne – ou l'équivalent pour les animaux non-humains. La conscience phénoménale ne serait donc le propre que d'un organe ; aucun être vivant biologiquement individué ne la posséderait de par sa nature. Même la philosophie mécaniste de Descartes n'abaissait pas la nature animale à de tels bas-fonds : quoique Descartes traitait effectivement les animaux comme des machines à réflexes incapables de pensée, il croyait tout de même qu'une telle explication était adéquate pour expliquer leur capacité d'éprouver des sensations, qu'il rattachait aux mouvements naturels de leur corps (Descartes 1953 : 127 ; Allen & Trestman 2017).

Comme le montre cette expérience de pensée, les conceptions que Parfit et Olson ont de nous-mêmes s'étendent à l'entièreté des êtres conscients ; nous, avec nos facultés mentales supplémentaires qui font de nous des personnes, n'en sommes qu'un cas particulier. Mais est-ce une manière plausible de comprendre le monde animal ? L'application du cas de transplantation cérébrale aux animaux non-humains expose un dilemme : soit tous les animaux cérébrés et conscients vont avec leur cerveau comme le propose l'animalisme compatibiliste ; soit aucun animal n'est un sujet d'expérience à proprement parler, capable d'éprouver sensations et émotions, puisque ce serait alors une autre entité, numériquement distincte de l'animal, qui aurait ces facultés. Mais si tous les êtres conscients que nous croyons être des animaux allaient là où va leur cerveau, pourquoi avancer alors qu'ils ne seraient pas réellement des animaux mais seulement des parties d'animal, plutôt que d'affirmer que ce qui va avec le cerveau, c'est justement l'animal lui-même ?

L'argument habituellement offert pour se considérer comme des êtres distincts des

(autres) animaux se fait sur la base de facultés mentales spéciales que nous seuls avons, celles-là mêmes qui font de nous des personnes. En vertu de ces facultés, nous aurions des conditions de persistance qui nous seraient propres, psychologiques plutôt que biologiques. Mais si les « animaux » conscients étaient eux-mêmes distincts de leur organisme comme l'impliquent les positions de Parfit et d'Olson – autrement dit, si être un animal conscient, *c'est* être une partie pensante d'animal, si c'est effectivement la règle – alors la raison d'affirmer que nous ne sommes pas des animaux mais une partie de l'animal ne s'avérerait qu'une manière déguisée d'affirmer ce que nous avons en commun avec tous les animaux conscients, c'est-à-dire d'exprimer que nous *sommes* de tels animaux. Nous humains ne serions pas un cas spécial. Affirmer que nous sommes l'organe ou la partie consciente d'un animal comme le fait Parfit reviendrait à dire que nous sommes des animaux conscients. Bref, lorsqu'on considère les facultés que nous partageons avec une partie du monde animal, notamment la conscience phénoménale, il peut paraître plus plausible d'avancer que de tels animaux, qui sont des sujets d'expérience, vont là où va leur cerveau, animaux humains y compris. Ce serait alors notre nature en tant qu'animaux conscients, plutôt qu'en tant que personne, qui ferait en sorte que notre cerveau est la partie essentielle de notre organisme.

Cela n'est certes qu'une lecture possible de cette expérience de pensée, et de telles expériences ne nous permettent que de comparer entre elles des intuitions. Si elle montre quelque chose, c'est que ce que nous disons à propos de nous-mêmes dans de tels cas devrait valoir pour les autres êtres conscients. Pour affirmer plus et appuyer l'idée que la conscience phénoménale serait essentielle à la survie des êtres ayant cette faculté, il faudrait des données empiriques allant en ce sens. L'étude d'une telle question pouvant faire l'objet d'un mémoire entier, nous allons simplement tenter de donner un aperçu de la manière dont il est possible de saisir le rôle central de la conscience pour l'organisme.

Une partie des efforts de recherche à ce sujet se centre sur la conscience chez les animaux non-humains, afin de mieux en saisir les bases évolutives et les raisons de son apparition. Cependant, l'attitude traditionnelle, béhavioriste, veut que la science ne pourrait donner réponse à des questions sur l'hypothétique vie subjective des animaux, puisque nous n'avons pas d'accès empirique direct à celle-ci. Attribuer la conscience phénoménale à des êtres sur la base de similarités entre leur comportement observable et

le nôtre, ainsi que d'homologies entre notre système nerveux et le leur, serait faire preuve d'un anthropomorphisme contestable. Les chercheurs dans ce domaine soutiennent par contre que la science a souvent étudié par des méthodes indirectes des processus ne pouvant être directement observés, notamment en physique (Griffin 2000 : 890 ; Panksepp 2011 : 1793).

Jaak Panksepp, pionnier dans la recherche sur les mécanismes neurologiques de l'émotion chez les animaux, y ayant travaillé depuis les années 60 jusqu'à sa mort l'an dernier, rapporte que les éléments de preuve les plus fructueux dans son champ de recherche ont été obtenus grâce à des stimulations électriques et chimiques de régions du cerveau. De telles stimulations provoquent des comportements émotifs chez les animaux, variables selon la région stimulée, et ces résultats servent à construire des hypothèses et modèles falsifiables, qui prédisent des comportements qui peuvent être vérifiés par la suite chez d'autres espèces, notamment chez les humains qui peuvent en rendre compte verbalement (Panksepp 2011 : 1793-4). De plus, les animaux eux-mêmes peuvent être entraînés à fournir un rapport comportemental sur leurs expériences subjectives. En voici un exemple :

« [D]irect stimulations of certain brain circuits can serve as rewards and punishments in various simple learning tasks. When given control of the stimulation—to turn it on or off—animals clearly indicate that they either like or dislike such forms of brain stimulation [...] [They] rapidly learn to turn off ESB that evokes fear- and anger-type responses, and they turn on brain stimulation that generates playful sounds, exploration, sexual eagerness and maternal care. » (Panksepp 2011 : 1794)

De tels résultats seraient bien difficiles à expliquer si ces animaux n'avaient pas d'expériences subjectives qui motivent leur agir. Il semble donc que la conscience animale ne soit pas un domaine fermé à la science.

En ce à trait aux émotions, Panksepp rapporte que les recherches depuis les années 50 confirment l'idée que ce sont des régions sous-corticales primitives qui en sont la source, et que ces régions seraient présentes chez tous les mammifères, sous une forme anatomiquement et neurochimiquement similaire :

« It has been clear for a long time that the most powerful affective experiences in humans are evoked by stimulating deep subcortical regions of the brain from where emotional behaviors can be evoked in animals. » (Panksepp 2011 : 1795)

Les recherches du neurologue Björn Merker montrent d'ailleurs que les enfants humains nés sans cortex cérébral sont conscients de leur environnement et capables d'éprouver des émotions (Merker 2007 : 78-80). Par conséquent, l'idée que notre vie mentale – cognitive *et* émotive – serait préservée par une transplantation de nos hémisphères cérébraux seuls serait erronée. Le cortex cérébral ne serait ni ce qui produit les émotions, ni, comme le croyait William James, ce qui produit l'expérience émotive, qu'il considérait n'être qu'une perception consciente de changements physiologiques (James 1884 : 203). Ce serait alors, en plus des recherches de Pinto, une raison supplémentaire de rejeter la *Cerebrum Transplant Intuition*.

Chez les mammifères non-humains qui ont été étudiés en détail – notamment divers primates, chats, chiens et rats – Panksepp avance que les données supportent l'existence d'au moins sept formes distinctes d'excitation émotive primordiales (exploration, peur, rage, jeu, désir sexuel, panique/deuil, souci parental – c'est moi qui traduis), et que ceux-ci jouent un rôle essentiel dans la survie de ces organismes. Ce sont ce qu'il nomme des « outils pour vivre » (Panksepp 2011 : 1802) qui non seulement informent les animaux de leur situation et de leur santé ; mais qui leur permettent de naviguer leur monde le plus sécuritairement possible ; et qui jouent un rôle central dans leur apprentissage comportemental :

« The primal affects are intrinsic brain value systems that unconditionally and automatically inform animals how they are faring in survival. They serve an essential function in emotional learning. The positive affects index “comfort zones” that support survival, while negative affects inform animals of circumstances that may impair survival. (Panksepp 2011 : 1791) [T]hey anticipate the kinds of survival needs that all organisms require to successfully navigate the world. » (Panksepp 2011 : 1798)

Le rôle des émotions est si important que Panksepp suggère qu'elles seraient fondamentales à la cohérence interne de l'organisme, de ce que je traduis par le « soi animal primaire » (*primary-process core-self*) :

« [I]f one simply considers the troublesome concept of “self”, it may need, at a primary-process level, to be anchored in the bodily emotional action coherences that characterize all organisms. » (Panksepp 2011 : 1801)

S'il n'a pas tort, il n'y aurait pas chez les êtres humains d'identité animale ni personnelle sans le noyau affectif sous-jacent à leurs capacités comportementales primaires, localisé dans le cerveau. Cela suggérerait que l'animal et la personne iraient tous deux là où va

leur cerveau.

Les recherches de Michel Cabanac et ses collègues, qui s'intéressent à l'apparition de la conscience dans l'arbre phylogénique, vont dans le même sens que celles de Panksepp, attribuant à la conscience phénoménale un rôle essentiel à la survie de l'organisme. Leurs résultats indiquent qu'il y a eu une transformation majeure lors du passage de la vie aquatique et amphibienne à la vie exclusivement terrestre, à la fois d'un point de vue de l'anatomie et de la chimie du cerveau que des comportements observables associés à la conscience chez l'humain (Cabanac 2009 : 268-9). Cela conduit Cabanac à suggérer que c'est la complexité de la vie terrestre qui a provoqué l'apparition de la conscience phénoménale – ou du moins son développement en une caractéristique centrale dans la constitution des organismes animaux. Son rôle fondamental serait de fournir aux animaux une manière flexible et efficace de déterminer le comportement optimal qu'ils devraient adopter, en l'occurrence celui qui maximise le plaisir et minimise la douleur :

« Consciousness may have emerged because of the increasing complexity of life in a terrestrial environment. In this new adaptive landscape, existence required more and more stimulus–response pathways; eventually, a point was reached where it became more efficient, in terms of speed and flexibility, to route all decision making through a single mental space. Within this space, different possible responses would be compared and judged according to the degree of pleasure they evoked, the aim being to maximize pleasure and to minimize displeasure. » (Cabanac 2009 : 271-2)

Une telle thèse, si elle n'est pas erronée, montre comment la conscience phénoménale pourrait être conçue comme ayant un rôle essentiel pour les animaux doués de cette faculté, soit celui d'orienter leur agir en vue de favoriser leur survie. Si la conscience jouait effectivement ce rôle – comme le concluent Panksepp et Cabanac, tout comme plusieurs auteurs qui s'opposent aux conclusions de Cabanac en ce qui a trait au moment de l'apparition de la conscience (Allen & Trestman 2017) – il serait alors tout à fait plausible de croire que nous, les animaux humains, allons là où va notre cerveau.

Par suite, si les animaux conscients l'étaient essentiellement – autrement dit s'ils étaient essentiellement des sujets d'expérience, ayant une vie sensible et émotive centrale à leur existence – il serait alors bien difficile d'imaginer que la personne soit séparable de l'animal, puisque la personne est également conçue comme un sujet d'expérience et

dépend du même cerveau. L'animal, le sujet d'expérience et la personne iraient de pair. Les concepts de sujet d'expérience et de personne référerait alors à la même entité réelle, l'animal humain, quoique saisi par des facultés différentes : la conscience phénoménale dans le cas du sujet, et la conscience de soi et la rationalité dans le cas de la personne.

3.2.3 – La conscience et le critère d'identité biologique

Attardons-nous maintenant aux difficultés auxquelles fait face l'animalisme compatibiliste quant à la formulation de notre critère d'identité. Si la conscience phénoménale était effectivement une fonction vitale des organismes conscients, et expliquait qu'ils iraient là où va leur cerveau, encore faudrait-il montrer comment la continuité du cerveau et de la conscience pourraient être intégrés à un critère d'identité animale qui serait purement biologique. En effet, à la fois le fait que les organismes conscients commencent leur vie sans cerveau ni conscience ainsi que l'idée que la conscience est une faculté psychologique semblent disqualifier d'emblée l'animalisme compatibiliste. Voici une piste de réponse qui pourrait servir à maintenir sa plausibilité malgré de telles critiques, sur laquelle nous allons élaborer ici : les *capacités* à développer un cerveau et la conscience seraient présentes dès le début de l'existence des organismes conscients, et elles pourraient être considérées de nature biologique parce qu'elles ne mettent en jeu aucun *contenu* psychologique et aucune relation entre de tels contenus. Nous allons commencer en abordant le problème de l'absence de facultés mentales au début de notre existence, pour ensuite nous attarder à la formulation du critère d'identité animale, et finir en considérant les implications du critère proposé.

D'abord, l'*argument du fœtus* que développe Olson semble impliquer qu'il faudrait écarter de notre critère d'identité toute caractéristique qui ne serait pas présente au commencement de notre existence. Notre organisme pouvant de toute évidence exister sans les organes et les facultés que nous développons suite à notre conception, ceux-ci ne seraient alors pas essentiels à la continuité de notre organisme. C'est une des raisons qui disqualifie le critère psychologique pour Olson.

Par contre, Olson lui-même ne se tient pas aux contraintes qu'il pose. Pendant plus de vingt ans, il a défendu que la partie essentielle à la survie de l'animal humain était

son tronc cérébral, une structure pourtant absente au début de son existence. Ce n'est d'ailleurs pas par souci de cohérence avec l'argument du fœtus qu'il a récemment changé d'avis, mais parce qu'il a pris connaissance des recherches de Shewmon supportant l'idée que le tronc cérébral ne serait pas nécessaire au maintien des fonctions vitales de l'organisme. Il croit donc que le fait d'intégrer à notre critère d'identité biologique un organe (ou une partie d'organe) qui n'est pas encore présent au début de notre existence ne serait pas en soi problématique, mais pourrait au contraire être tout à fait cohérent s'il acquiert une fonction vitale. Nous pouvons alors réinterpréter son argument du fœtus, et juger qu'il vise l'approche psychologique non pas parce que les fœtus commencent leur existence sans traits ni facultés psychologiques, mais plutôt parce que cette approche propose un critère d'identité fondé sur des éléments non-vitaux à l'existence de notre organisme, en l'occurrence nos contenus mentaux et/ou les relations entre eux. Notons que par contenus mentaux nous entendons les expériences, états ou événements mentaux ; ceux-là mêmes qui selon Parfit et bien des lockéens sont les principaux constituants de sujets conscients.

De plus, Olson a tout aussi longtemps souscrit à la *Whole Brain Transplant Intuition*, l'idée que nous, l'organisme humain, irions là où va notre cerveau, et ce parce qu'il contient le tronc cérébral. Ainsi, l'idée que le cerveau (ou le tronc cérébral), un simple organe, *serait* l'organisme original sévèrement mutilé n'est pas non plus hors de question pour lui, tant qu'il est conçu comme ce qui assure les fonctions vitales principales de l'organisme. Par conséquent, puisque l'animalisme compatibiliste soutient que le cerveau remplit une fonction vitale essentielle en supportant la conscience phénoménale, il est cohérent pour cette position de considérer qu'il serait l'organisme original lorsque séparé du reste du corps.

Si un compatibiliste peut alors avancer que le critère d'identité des êtres humains et des autres animaux conscients consiste en la continuité de leur cerveau et de sa capacité à supporter la conscience, il lui faut tout de même énoncer ce critère de façon convenable afin qu'il soit valable durant toute leur existence, même durant la phase fœtale. Nous avons suggéré plus tôt, en présentant la seconde position de Parfit, que formuler le critère en terme d'une *capacité* à développer un cerveau capable de supporter la conscience serait un moyen propice à cette fin (2.2.4). Une première version du critère

de notre identité animale requerrait alors (1) la continuité de suffisamment de notre cerveau pour maintenir la capacité à supporter la conscience, *ou* (2) la continuité de notre capacité à développer un cerveau ayant la capacité à supporter la conscience. Un fœtus humain pourrait alors être considéré comme un animal humain à plein titre, et il en vaudrait de même pour les autres organismes essentiellement conscients.

Une telle formulation est-elle toutefois appropriée ? Ou faudrait-il plutôt séparer l'exigence de la continuité du cerveau de celle de la continuité de la capacité à supporter la conscience ? Deux auteurs qui conçoivent notre identité en terme de l'identité de conscience, Peter Unger¹⁵ et Jeff McMahan, proposent tous deux un critère nécessitant seulement la continuité des réalisateurs physiques de la conscience (Unger 1990 : 109 ; McMahan 2002 : 67). Quoiqu'ils parlent du critère d'identité *personnelle* plutôt que de celui de l'organisme humain, leurs apports demeurent pertinents ici.

La raison pour laquelle ces auteurs croient que la continuité des réalisateurs physiques de la conscience seraient suffisants, c'est que leurs critères ne requièrent que la continuité de la *capacité* à être conscient, plutôt que la continuité de la conscience elle-même comme le défendait Locke. La continuité de contenus psychologiques et/ou des relations entre eux nécessiterait un critère psychologique, alors que celle de capacités psychologiques pourrait se contenter d'un critère de même nature que les réalisateurs de ces capacités. Cela permettrait d'ailleurs à des critères ainsi construits d'être compatibles avec des périodes d'inconscience, lors du sommeil par exemple, ou encore avec la perte de mémoire et des maladies tel l'Alzheimer (McMahan 2002 : 67-8). Même si ma conscience était temporairement interrompue, ma capacité à être conscient demeurerait intacte. Par contre, alors que McMahan identifie les réalisateurs de cette capacité au cerveau, Unger laisse place à ce qu'ils soient hypothétiquement remplaçables (Unger 1990 : 106). Nous allons faire abstraction de cette possibilité ici, mais en glisserons un mot dans la prochaine sous-section.

Cela dit, McMahan insiste sur le fait que malgré que son critère n'invoque que la continuité du cerveau pour rendre compte de la capacité à être conscient, ce critère n'est pas simplement *physique*. Il n'exige pas seulement la continuité de la matière qui compose le cerveau ou le remplacement graduel de celle-ci, mais requiert également une

¹⁵ Nous avons parlé d'Unger plus tôt (1.2.1), en référant à son texte de 1979 où il défendait un éliminativisme. Il a changé de position depuis.

continuité *fonctionnelle* du cerveau, qui préserve les capacités psychologiques de base (McMahan 2002 : 68-9). Unger abonde dans le même sens, quoiqu'il croit qu'il serait plus approprié de dire que de telles capacités seraient maintenues par la continuité d'une *structure* physique adéquate (Unger 1990 : 105). Unger précise que ce n'est toutefois pas la continuité du *fonctionnement* du cerveau dans tous ses états – physiques et mentaux – qui serait requise, contrairement à ce qu'avance Nagel (1986 : 40-1). Il nous invite à imaginer un processus de congélation durant lequel le cerveau d'un individu cesserait temporairement de fonctionner mais conserverait ses capacités, retrouvant sa fonctionnalité après la décongélation (Unger 1990 : 105). Il réitère que ce n'est que la préservation des capacités qui importe :

« For survival, there need only be continuity of certain psychological *capacities*; there need *not* be continuity of any psychological *activities*, whether conscious or non-conscious. » (Unger 1990 : 105)

Bref, il semble que même si le critère d'identité animale en terme de continuité du cerveau ne nécessiterait pas de clause supplémentaire afin d'y inclure la capacité à supporter la conscience, il faudrait tout de même préciser ce qu'on veut dire par une telle continuité pour la distinguer d'une simple continuité physique ou d'une continuité de son activité. Nous laissons cette discussion en suspens ici.

Olson, quant à lui, s'oppose vivement à cette approche par capacités, croyant qu'elle ne fait que réintroduire l'approche psychologique par la porte de derrière. Dire qu'un fœtus humain a une capacité à développer des capacités mentales – autrement dit qu'il a des capacités mentales de deuxième ordre – ce ne serait qu'un truc pour étirer indûment le critère de continuité psychologique jusqu'au début de l'existence de l'organisme afin d'éviter l'argument du fœtus (Olson 1997 : 85-7). Cela permettrait à des lockéens de dire que nous sommes essentiellement des personnes tout en ayant été des fœtus. Olson trouve que cette manière d'avancer que nous serions en « continuité psychologique » avec un fœtus qui n'avait aucun trait psychologique actualisé à cette étape de son développement est tout à fait absurde.

L'animalisme compatibiliste ne cherche toutefois pas à réintroduire une approche psychologique de l'identité ni à affirmer que les fœtus seraient des personnes, proposant plutôt que la capacité à être conscient serait essentielle à l'identité des *animaux* conscients. Cette position ne serait pas mieux exprimée sous forme d'un critère

psychologique de l'identité de ces animaux, et ce même si elle ne faisait pas appel à des contenus mentaux, parce qu'elle impliquerait une rupture radicale dans le monde animal. Il y aurait alors deux groupes d'animaux : ceux qui satisfont au critère biologique mais ne possèdent aucune capacité mentale, et ceux qui satisfont au critère psychologique en vertu de telles capacités. Cela reviendrait à rejeter la continuité évolutive des espèces naturelles et irait à l'encontre de l'objectif d'affirmer que les êtres humains sont essentiellement des animaux plutôt que des personnes.

Selon Olson par contre, le simple fait de postuler que les conditions de persistance des animaux conscients seraient différentes de celles des autres animaux revient à postuler une telle rupture. Les animaux n'auraient pas alors des conditions de persistance communes, en vertu de leur appartenance au sortail de substance *animal*, mais plutôt en vertu de leurs propriétés spécifiques comme nous l'avons expliqué plus tôt (2.1.3). Wiggins s'oppose par contre à cette conception et propose que les sortails de substance des animaux qui déterminent leurs conditions de persistance sont le nom des espèces auxquelles ils appartiennent. Les animaux auraient alors en commun un critère d'identité *biologique*, quoique ce qui est vital à la continuité de leur organisme dépendrait de la sorte d'organisme qu'ils sont. D'ailleurs, Olson lui-même n'accordait-il pas aux animaux humains des conditions de persistance qui dépendaient de leur tronc cérébral, alors que bien des animaux n'ont même pas cette structure ? Une approche wigginsienne semble être la meilleure pour l'animalisme compatibiliste ; elle permet d'affirmer que la conscience n'est vitale que pour certaines espèces animales, sans toutefois nécessiter un critère radicalement différent, non-biologique.

Cette position a cependant une conséquence qui pourrait en rebuter plusieurs. Si la conscience était effectivement une faculté essentielle des organismes qui la possèdent, elle devrait alors être comprise comme une capacité *biologique* – du moins chez les êtres conscients connus à ce jour. Tout comme c'est un fait biologique qu'un fœtus a la capacité de développer la faculté de vision, ce serait un fait biologique qu'il ait la capacité de développer la faculté d'être conscient. Nagel définit d'ailleurs la conscience comme un phénomène qui n'existe que *pour* un organisme, comme on l'a vu. Cela viendrait également rejoindre McMahan et Unger dans leur volonté de considérer la continuité des réalisateurs de la conscience comme un critère qui n'est pas simplement physique, mais

qui requiert quelque chose de plus, une sorte de structure fonctionnelle capable d'assurer la conscience et sa dimension subjective. Si la capacité à être conscient était effectivement de nature biologique plutôt que purement physique ou psychologique, un critère biologique serait alors tout indiqué.

Une telle thèse, si elle était vraie, répondrait d'ailleurs à la position de Shoemaker selon laquelle des êtres ayant des facultés mentales ne pourraient être des animaux parce qu'ils ont des conditions de persistance mutuellement exclusives. Rappelons que, selon lui, avoir des états mentaux signifie, en raison de leur rôle fonctionnel, que la série d'états mentaux qui en découle appartient au même individu qui possède ces états initiaux (2.3.1 ; voir aussi Shoemaker 1997 : 295-6). Le fait d'avoir des états mentaux déterminerait donc les conditions de persistance de l'individu qui les a. Il en irait de même pour les états biologiques. Par suite, les êtres ayant des états mentaux auraient des conditions de persistance psychologiques, alors que les êtres ayant des états biologiques auraient des conditions de persistance biologiques. Aucun animal ne pourrait alors être conscient ni avoir la faculté de penser. Cependant, si la conscience était une faculté *biologique* plutôt que purement psychologique, les dimensions biologique et psychologique d'un organisme conscient ne seraient pas mutuellement exclusives, mais feraient partie d'un tout fonctionnel. Une entité ayant des états mentaux pourrait alors avoir des conditions de persistance biologiques. Des animaux comme nous pourraient alors être conscients et avoir la faculté de penser.

L'animalisme compatibiliste ainsi formulé a une autre conséquence : si nous – et tout animal *essentiellement* conscient – perdions de façon permanente notre capacité à être conscient ou notre capacité à développer cette capacité, nous serions tout simplement morts. Nous ne pourrions alors tomber dans un état végétatif – à moins qu'il ne soit que temporaire, auquel cas nous aurions une situation analogue à celle de la congélation de l'organisme qu'imaginait Unger, interrompant nos activités conscientes sans toutefois porter atteinte à nos capacités. Olson juge que c'est une erreur de chercher une position qui aurait une telle conséquence, parce qu'elle implique qu'une simple perte de capacités voudrait dire qu'un animal meurt et qu'un autre de nature différente prendrait sa place (Olson 1997 : 88-9). Un compatibiliste qui considère vitale la capacité à être conscient affirmerait plutôt qu'un animal ayant perdu cette capacité serait *biologiquement* mort, ne

laissant derrière lui que du tissu vivant plutôt qu'un organisme.

Olson objecte qu'on pourrait imaginer une technique pour régénérer les réalisateurs de la conscience, de manière à ce qu'un tel corps retrouve « sa » capacité à être conscient (Olson 1997 : 87-8). Ce corps aurait ainsi conservé la capacité à *développer* la capacité à être conscient durant son épisode végétatif, tel un fœtus durant son développement ; il serait donc le même organisme selon le critère compatibiliste. La perte de la capacité à être conscient ne signifierait donc pas la mort biologique de l'animal. À cette objection, un compatibiliste devrait répondre que contrairement à un fœtus, un corps adulte n'a plus la capacité à développer *par lui-même* la capacité à être conscient. Régénérer artificiellement cette capacité créerait donc un *nouvel* animal avec le tissu de celui qui est mort, plutôt que de la redonner à l'animal original. La perte de cette capacité signifierait donc notre mort définitive.

En somme, afin que l'animalisme compatibiliste soit plausible, il faudrait considérer que la capacité à être conscient est déjà présente chez un fœtus humain, comme capacité de second ordre qu'il peut actualiser durant son développement. Cela implique qu'une telle capacité serait biologique, ainsi que vitale et essentielle puisqu'elle déterminerait notre critère d'identité. La perte de cette faculté, normalement associée à la mort cérébrale, serait alors identifiée à la mort de l'organisme. Si tout cela était faux, la position serait alors à revoir dans son intégralité, voire à rejeter.

3.2.4 – La conscience, les contenus mentaux, et ce qui importe

Jusqu'ici, nous avons fait un survol général de l'animalisme compatibiliste et de ses implications. Nous avons insisté sur le fait qu'on pouvait considérer des capacités « psychologiques » – telles la sensibilité, l'émotivité, et plus largement la conscience phénoménale – comme essentielles à un organisme biologique, tant qu'on les distingue des contenus mentaux. Cela expliquerait que la continuité du cerveau soit requise pour le maintien de l'identité des animaux conscients.

Nous allons maintenant tester les limites de cette position à l'aide de deux expériences de pensée, la première séparant le sujet conscient de ses contenus mentaux, la seconde le sujet conscient de son cerveau. De telles expériences résultent en des intuitions conflictuelles quant à quel élément préserverait notre identité – nos contenus

mentaux, notre conscience, ou notre cerveau – et menacent de remettre en question le compatibilisme. Nous allons tenter de montrer leurs implications et la manière dont un compatibiliste pourrait y répondre afin de préserver sa position.

Commençons par considérer le scénario où l'on sépare l'animal conscient de ses contenus (ou états) mentaux. Nous avons avancé plus tôt que l'animalisme compatibiliste implique que la personne serait inséparable de l'animal, parce qu'elle s'y réduit. Tous deux iraient d'ailleurs là où va le cerveau de l'animal, puisqu'ils dépendent de sa capacité à être conscient. Mais cette inséparabilité est-elle justifiée ? Si les contenus de conscience étaient séparables du sujet conscient, cela reviendrait-il à dire que la personne serait séparable de l'animal, et, par suite, que l'animalisme compatibiliste serait faux ou à réviser dans son intégralité ?

Nous allons examiner cette possibilité à l'aide de la procédure hypothétique qu'envisage Shoemaker, qu'il nomme *Brain State Transfer* et que nous avons vue plus tôt (2.3.1). Cette procédure copie l'état des réalisateurs physiques des états mentaux d'un individu à un autre cerveau qui dispose des capacités mentales nécessaires. Imaginez qu'on vous fasse une telle procédure, transférant ainsi vos contenus de conscience à un clone fraîchement sorti de l'usine, mais que cela ait l'effet malencontreux d'effacer ces contenus de votre cerveau. L'original aurait donc conservé votre conscience, alors que le clone aurait une copie de vos états mentaux. Selon la version libérale du critère psychologique de l'identité personnelle, qui ne requiert que continuité et connectivité psychologiques peu importe la manière dont ils sont préservés, le clone serait la personne que vous étiez avant la procédure. Tous s'entendraient pour dire que l'original est le même animal. Lequel seriez-vous alors ?

Shoemaker lui-même est ambivalent quant à l'issue de la procédure, parce qu'elle viole l'idée que nous sommes ce qu'il nomme des « relatively autonomous self-perpetuators », c'est-à-dire des êtres dont la persistance est assurée par une causalité qui leur est immanente plutôt qu'externe (Shoemaker 1997 : 290). Néanmoins, si on considérait que le clone serait la personne que nous étions avant la procédure, cela voudrait dire que la personne serait séparable de l'animal et ne s'y réduirait donc pas, contrairement à ce qu'avance l'animalisme compatibiliste. Cela aurait pour conséquence que dans notre corps logeraient deux entités distinctes, avec tous les problèmes que ça

comporte. Mais est-ce une conséquence nécessaire ? Est-il possible qu'une entité ontologiquement réduite survive séparément de sa base de réduction ? Si c'était possible, cela remettrait en question les assumptions qui nous ont conduit à la position compatibiliste, en particulier l'idée que *nous* devrions être une entité réelle afin d'être des penseurs et des agents. Ce serait une piste à examiner davantage, qu'on veuille défendre la position ou la rejeter.

Un compatibiliste voulant préserver sa position devrait traiter un tel cas de la manière suivante. Il devrait considérer que la procédure de transfert de contenus de conscience ne pourrait préserver l'identité personnelle, mais copierait seulement un ensemble de propriétés qui entrent dans sa constitution. Dire qu'il y aurait « identité » personnelle serait alors une méprise langagière. L'individu résultant du transfert, malgré qu'il aurait nos contenus mentaux, serait tout simplement un autre individu, et donc une autre personne. Nous, l'animal humain, serions numériquement le même indépendamment de nos contenus mentaux.

Supposons à présent que le clone mental n'est ni la personne ni l'animal original, comme le suppose toute position animaliste, tout comme les positions psychologues qui requièrent la continuité du cerveau défendues par Parfit. Il demeure possible de se questionner sur ce qui importe le plus pour nous, entre notre conscience et nos contenus mentaux. Si on était forcé de ne préserver que l'un des individus sur lesquels la procédure a été effectuée, l'animal ayant perdu la mémoire ou le clone ayant nos contenus mentaux, lequel choisirions-nous ? Parfit a fameusement soutenu que notre identité en tant que sujet importe peu, et que ce sont plutôt les relations psychologiques constituant notre vie mentale dont nous nous soucions. Selon lui, même si l'identité était perdue lors du transfert, ce qui importe serait tout de même préservé chez le clone ayant nos contenus mentaux. Olson est d'accord avec ce point de vue, tout comme Shoemaker, malgré que ce dernier croit qu'une position qui dissocie notre identité de ce qui importe pour nous serait instable et insatisfaisante (Shoemaker 1997 : 299). Que répondrait l'animalisme compatibiliste ?

Rappelons que nous avons avancé dans le premier chapitre que l'animalisme pourrait être une version plausible de la *ego theory*, qui saurait supporter l'idée que c'est notre identité qui importe le plus, quoique nos contenus mentaux aient une valeur non-

négligeable (1.1.4 ; 1.2.3). L'animalisme compatibiliste serait-il capable de défendre une telle position ?

Les recherches de Panksepp et Cabanac sur le rôle évolutif de la conscience phénoménale et des émotions pourraient nous donner une piste de réponse. Si, comme ces auteurs l'avancent, la conscience était avant tout un système de valeur dont le rôle serait de guider notre agir pour favoriser notre survie, cela signifierait que notre conscience est au cœur de tout souci de soi, comme l'affirmait d'ailleurs Locke (1975 : II.XXVII, §17). Par l'entremise d'affects tels la douleur et le plaisir, elle serait à l'origine de toute attribution de valeur à des situations, des activités, et des contenus mentaux. Mes contenus mentaux n'auraient alors d'importance pour moi que parce que je leur ai conféré une forte charge émotive. Indépendamment de ma conscience qui évalue des contenus, un ensemble de contenus particuliers n'aurait aucune valeur intrinsèque pour moi. Sur cette base, il serait possible d'avancer que notre conscience, en tant que source de toute valeur pour nous, importe bien plus que ce qu'elle a valorisé. Et puisque selon l'animalisme compatibiliste la conscience est essentielle à notre identité, il serait possible de défendre que c'est notre identité qui importe le plus pour nous, contrairement à ce que défendent Parfit et Olson. Il y a d'ailleurs fort à parier que notre système de valeur conscient ne soit que la concrétisation d'une valeur implicite à la vie de tout individu. Chaque expérience de douleur et de plaisir exprimerait avant tout l'importance de l'organisme pour lui-même.

Terminons en examinant un second scénario qui semble incompatible avec l'animalisme compatibiliste : la séparabilité de la conscience (et ses contenus) d'avec le cerveau. L'identité animale telle que conçue par l'animalisme compatibiliste semble nécessiter les deux composantes, puisque c'est le cerveau qui selon cette position réalise physiquement la capacité à être conscient ainsi que l'activité consciente. Mais s'il était possible que notre conscience, et non pas seulement nos contenus mentaux, soit transférée à d'autres réalisateurs physiques, cela reviendrait-il à tuer l'animal tout en préservant la personne ? Comme dans le cas de l'expérience de pensée précédente, cela voudrait dire que la personne ne se réduirait pas à l'animal, et mettrait la viabilité de l'animalisme compatibiliste en doute.

Un transfert de notre conscience à un clone de notre cerveau ou à un autre

cerveau humain serait ce qu'il y a de plus plausible à imaginer. Comme l'avance Nagel, notre forme particulière de conscience est spécifique à notre espèce, dépendante du type d'organisme que nous sommes (Nagel 1974 : 174). La structure de notre cerveau, qui réalise physiquement notre conscience, en atteste. Notamment, celui-ci est constitué d'aires sensorielles qui correspondent topographiquement aux surfaces sensibles de notre corps, comme le montre l'homunculus de Penfield (voir l'annexe A). Plus la sensibilité d'une surface du corps est fine, plus la surface du cerveau qui lui est dédiée est grande, en particulier dans le cas des mains et de la bouche. Un substitut à notre cerveau devrait donc reproduire une telle structure, ou du moins son effet qualitatif.

Par contre, il n'est pas clair qu'on puisse distinguer conceptuellement un transfert de notre conscience et nos contenus mentaux d'un simple transfert de nos contenus, sans données empiriques. D'ailleurs, une telle procédure irait à l'encontre de l'intuition que nous sommes des êtres qui sont cause de leur propre persistance, plutôt que de persister en vertu d'une intervention externe et artificielle, comme l'a souligné Shoemaker. Unger propose une expérience de pensée qui serait compatible avec cette intuition, qu'il est utile d'adapter ici (Unger 1990 : 107). Imaginons que notre cerveau soit graduellement remplacé par des structures de silicone, qui remplissent les fonctions des parties qu'elles remplacent. Le processus alloue un temps suffisant pour que notre conscience fasse siennes ces structures, tout comme elle ferait siennes de nouvelles cellules qui viennent remplacer les anciennes durant le fonctionnement normal de notre cerveau. À la fin de l'opération, notre cerveau serait entièrement remplacé par un cerveau de nature différente. Notre conscience serait-elle être préservée par un tel processus, s'il était exécuté de façon appropriée ?

Il y a bien de bonnes raisons d'en douter, mais supposons tout de même que ce soit possible. L'opération ayant vraisemblablement éliminé l'animal puisque son cerveau lui était essentiel, alors même que sa conscience est préservée, cela signifierait-il que l'animalisme compatibiliste est à rejeter ? Un compatibiliste devrait-il alors adopter une approche psychologique de l'identité, et concéder que nous sommes essentiellement des personnes ? Unger avance qu'il serait très difficile de réfuter l'idée que le cerveau de silicone ne serait pas le cerveau de l'animal lui-même. Il serait en effet tout à fait possible d'avancer que l'opération aurait l'effet de *transformer* son cerveau, plutôt que de le

remplacer (Unger 1990 : 107). D'ailleurs, si la conscience était effectivement un résultat de l'évolution *biologique*, et était essentielle à l'identité des organismes conscients, il serait même possible d'avancer que tout ce qui préservait cette conscience préserverait en fait l'organisme lui-même. John McDowell estime que lorsqu'on se distancie de la division cartésienne entre corps et conscience, le prolongement de la carrière d'une conscience humaine apparaît alors comme la continuation de la même vie animale (McDowell 1997 : 237). L'identité de la conscience, indépendamment de la nature de son support, ne consisterait pas en un critère psychologique, mais biologique. Si une telle réponse peut sembler abracadabrante, elle ne fait que correspondre à la nature de l'expérience de pensée considérée ici.

Bref, il y aurait moyen pour un animaliste compatibiliste de maintenir sa position face à des expériences de pensée qui tentent de la mettre à mal, et d'appuyer l'idée que la personne ne pourrait exister indépendamment de l'animal auquel elle se réduit. Cela implique de minimiser l'importance des contenus de conscience ainsi que du cerveau en tant que réalisateur de la conscience animale. Ce faisant, il serait alors possible de défendre que c'est de notre identité dont nous nous soucions le plus, parce qu'elle serait source de toute valeur consciente pour nous, ainsi que le moyen par lequel l'organisme manifeste sa propre valeur pour lui-même. Par contre, si des entités ontologiquement réduites pouvaient survivre à la mort de leur base de réduction, tous les arguments énoncés jusqu'ici seraient à revoir. Nous pourrions alors être une entité conceptuelle, la personne, « capable » de sauter de corps en corps par le seul transfert de nos contenus mentaux. Les entités ayant le plus d'impact sur le réel pourraient alors n'avoir aucune réalité propre. Il va sans dire qu'une telle position nous semble totalement absurde.

*

Revenons sur les points principaux de l'animalisme compatibiliste dont nous venons de donner un aperçu général. Nous avons développé plusieurs raisons de croire que nous sommes des animaux humains plutôt qu'une autre sorte d'entité. Premièrement, il semble que nous devons être des entités *réelles* afin de pouvoir se comprendre comme des *sujets d'expérience*, des *penseurs* et des *agents* non-dérivatifs, c'est-à-dire qui ressentent, pensent et agissent par eux-mêmes. Deuxièmement, nous devons être des entités *matérielles* afin de nous concevoir comme faisant partie intégrante du monde

naturel, physique, et inscrits dans un univers spatio-temporel. Et troisièmement, nous devons être des entités de nature *biologique* afin de nous comprendre comme participants à l'évolution des espèces naturelles, et de rendre compte adéquatement du lien de parenté que nous avons avec les autres animaux, à la fois au niveau de notre constitution que du noyau émotif et motivationnel à la base de notre comportement. Seule l'affirmation que nous sommes numériquement identiques à des animaux humains, et que les personnes s'y réduisent pleinement, semblerait être capable de satisfaire ces exigences.

Ensuite, nous avons avancé que pour qu'une telle conception de nous-mêmes comme *une* entité soit cohérente, il faudrait défendre que les animaux conscients, comme nous le sommes nous-mêmes, sont essentiellement leur cerveau, et que sans ce cerveau, leur corps ne pourrait être considéré comme un organisme, même s'il continue à assurer certaines fonctions biologiques. Une première raison pour cela réside dans la *Transplant Intuition* qui est d'une force intuitive considérable. Une seconde provient de la constatation que toute autre position impliquerait que la séparation du cerveau d'avec le corps produirait soit une *personne-résiduelle*, soit un *animal-résiduel*, et impliquerait alors que tout animal conscient serait en fait deux entités numériquement distinctes. Si nous pouvions survivre séparément du support de notre conscience, notre nature ontologique et par suite notre compréhension de nous-mêmes serait irrémédiablement fragmentée.

Nous avons alors expliqué qu'afin de justifier que nous allons là où va notre cerveau, il faudrait pouvoir montrer que ce dernier remplit des fonctions vitales essentielles qui vont au-delà des fonctions auto-régulatrices de l'organisme, sans quoi le corps décérébré pourrait également être considéré comme l'organisme. Nous avons suggéré qu'en plus de l'intentionnalité, la conscience phénoménale serait une de ces fonctions essentielles parce qu'elle semble jouer un rôle central dans l'orientation de notre agir dans le but d'assurer notre survie. Nous avons par la suite proposé que notre critère d'identité (animale) serait formulé en termes de *capacités* à développer un cerveau et à être conscient, et ne ferait intervenir aucun contenu mental – quoique les contenus mentaux puissent faire partie du critère d'identité personnelle. Cela nous permettrait alors de considérer que nous avons été des fœtus, et que nous demeurons le même individu si nos contenus mentaux sont affectés ou si notre conscience est temporairement

interrompue. Un tel critère impliquerait alors que la conscience serait une faculté biologique, et que la perte de toute capacité à être conscient et à développer cette capacité entraînerait la mort.

Enfin, si la conscience était effectivement essentielle à notre identité animale, cela pourrait expliquer la croyance intuitive que c'est notre identité qui importe lorsqu'on se soucie de notre survie, conformément à la *ego theory*. En effet, la conscience serait le moyen par lequel toute chose acquiert de la valeur *pour nous*, manifestant par le fait même que ce sujet à qui la valeur des choses est subordonnée serait la valeur première pour lui-même.

CONCLUSION

Au tout début de ce mémoire, nous avons posé la question suivante : que sommes-nous ? C'est-à-dire quelle sorte d'entité sont les êtres humains ? Nous avons entrepris d'explorer des réponses modernes à cette question ainsi que les problèmes de chacune à travers le débat opposant Parfit aux animalistes classiques. La philosophie de Parfit a été particulièrement appropriée pour une telle investigation, en raison de son point de départ lockéen ainsi que des modifications successives de sa position en réaction aux arguments de ses opposants. Elle a tracé le chemin que nous avons suivi dans notre étude et a fait ressortir les arguments charnières sur lesquels prennent appui les diverses conceptions de notre nature ontologique.

Passons brièvement en revue les positions principales que nous avons examinées durant notre parcours afin d'en faire la synthèse. Nous avons avant tout souligné qu'il y a deux catégories générales sous lesquelles se classe toute conception de notre nature ontologique : la *ego theory* d'une part, non-réductionniste et conforme à nos intuitions, selon laquelle nous sommes une entité réelle, ayant une existence propre ; et la *bundle theory* d'autre part, réductionniste et d'inspiration humienne, selon laquelle nous n'avons pas d'existence en dehors des éléments qui entrent dans notre constitution.

Nous avons commencé notre étude par la position initiale de Parfit, qui s'inscrit dans le prolongement de l'approche psychologique de Locke. Or, le point central de Locke, c'est que notre conscience fait partie intégrante de ce que nous sommes, parce que nous sommes avant tout une entité qui se saisit comme *soi*, c'est-à-dire qui par sa conscience s'approprie de son histoire vécue et s'y identifie. Locke, et bien d'autres à sa suite, considèrent que cette entité essentiellement consciente de soi que nous sommes, à laquelle est aussi attribuée la pensée rationnelle, c'est ce qu'on nomme la *personne*. Nous avons également vu qu'à la fois Locke et Parfit emploient des expériences de pensée qui montrent que les connaissances qui leur sont contemporaines n'éliminent pas la possibilité que notre conscience persiste séparément d'une même entité réelle – fût-elle une substance matérielle, immatérielle, ou encore l'animal humain. Nous n'aurions

donc pas de raison suffisante d'identifier la personne à l'une de ces entités, si bien qu'elle est alors comprise comme une sorte d'entité numériquement distincte de toute autre, c'est-à-dire comme un sortat de substance. Puisque la personne n'est alors saisie que comme conscience, ses conditions de persistance sont psychologiques.

Là où les positions de Locke et Parfit divergent, c'est dans les conclusions qu'ils tirent de leur analyse. Alors que Locke faisait preuve de parcimonie et se retenait de préciser notre nature ontologique au-delà de sa dépendance à notre conscience, et ce malgré qu'il était intimement convaincu que la version cartésienne de la *ego theory* était vraie, Parfit fait un pas vers Hume et rejette explicitement l'existence des âmes ainsi que la possibilité que nous soyons une entité réelle, remaniant alors l'approche lockéenne en une version de la *bundle theory*. Comme nous l'avons vu, nous n'aurions pas selon lui d'existence propre ; plutôt, notre existence se réduirait à celle des éléments qui nous constituent, nos événements mentaux avant tout, et nous ne serions alors que le *bundle* conceptuel les réunissant. C'est là le réductionnisme constitutif de Parfit. Si nous existons en dépit de la réduction, ce n'est alors qu'en vertu de notre schème conceptuel qui requiert l'usage de sujets grammaticaux dans l'expression de certains énoncés. Ainsi, par le refus d'associer la conscience à une entité concrète, nous nous trouvons relégués hors de la nature et du réel, confinés au seul domaine du langage.

En réponse, nous avons suggéré que la conclusion de Parfit est trop hâtive, et que le rejet de l'existence des âmes n'implique pas nécessairement que notre conscience n'est pas dépendante d'une même entité réelle, à savoir un animal humain. Nous sommes donc allé voir les arguments des animalistes, qui soutiennent qu'en distinguant numériquement la personne de l'animal, les positions lockéennes telle celle de Parfit s'exposent à de sérieux problèmes. En effet, malgré que dans tous les cas réels chaque personne soit toujours liée au même animal humain, les positions lockéennes impliquent que nous, les personnes, ne sommes pas des animaux et n'en avons jamais été ; plutôt, nous n'aurions commencé à exister que lorsque ces animaux eurent développé leur faculté de penser. Nous avons vu que la manière la plus plausible d'accommoder la distinction lockéenne dans un monisme ontologique est le constitutionnalisme, selon lequel l'animal et la personne coïncident dans le même espace et partagent leur matière. Cependant, si le constitutionnalisme était vrai, les animaux humains auraient tout autant accès à leur

cerveau que les personnes, et pourraient alors penser et être eux-mêmes des personnes lockéennes, capables de conscience de soi et de pensée rationnelle. Il y aurait alors deux personnes là où l'on croit qu'il n'y en a qu'une, toutes deux pensant les mêmes pensées et posant les mêmes actions. Ainsi, non seulement les positions lockéennes non-dualistes dédoublent-elles les entités pensantes, mais elles conduisent à la surdétermination causale de toute activité humaine.

Nous avons alors examiné la solution des animalistes à ces problèmes, qui consiste à abandonner la distinction lockéenne à la source de ceux-ci. Selon l'animalisme, nous sommes chacun un animal humain, et la personne que nous sommes n'est pas une entité numériquement distincte de lui, mais est plutôt une manière de le décrire valable durant une partie de sa vie. Autrement dit, les personnes seraient un sortal de phase plutôt qu'une sorte d'entité fondamentale au même titre que les espèces animales. Par contre, cette solution vient au prix de l'intuition première de Locke : puisque les animaux humains existent durant une partie de leur vie sans facultés mentales ni contenus mentaux, l'animalisme classique écarte alors l'idée que notre conscience fait partie inhérente de notre nature ontologique et de notre identité. Nous avons également pu constater que l'animalisme est confronté au problème épineux de justifier pourquoi ce serait l'animal qui pense plutôt qu'une partie de lui tel le cerveau.

Nous nous sommes ensuite attardé à la réception de ces arguments par Parfit. Celui-ci concède que les animalistes ont mis le doigt sur des problèmes légitimes qui exigent une réponse, mais refuse de se défaire de l'intuition lockéenne qui accorde une place centrale à la conscience. Nous avons vu que sa première réaction consiste à se rallier aux animalistes quant à notre nature ontologique : nous serions donc essentiellement des animaux humains, les personnes n'étant qu'une manière de les saisir par leur capacité d'être conscients de soi et rationnels. La deuxième position de Parfit est donc une version de la *ego theory*. Par contre, elle se distingue de l'animalisme classique en ce qu'elle préserve l'approche lockéenne, psychologique, de l'identité personnelle, plutôt que d'affirmer que le critère d'identité de la personne serait tout simplement celui de l'animal. Par conséquent, afin de maintenir que la personne irait là où va notre cerveau et notre vie mentale, conformément à la *Transplant Intuition*, et d'éviter l'implication qu'elle ne serait pas un animal, Parfit soutient que les animaux humains sont

essentiellement leur cerveau plutôt que le reste de leur corps. Le cerveau serait essentiel à la survie de l'animal *et* de la personne, parce qu'il serait à la fois l'organe de contrôle de l'organisme et la source de la vie mentale de la personne.

Nous avons souligné que Parfit n'endosse toutefois pas toutes les implications de sa deuxième position, parce qu'il croit tout de même qu'une personne n'est pas toujours ontologiquement réductible à un animal humain – lors de la division des hémisphères cérébraux par exemple. C'est en partie pourquoi il développe par après une troisième position qui lui permet de répondre aux arguments animalistes tout en préservant l'idée que personne et animal sont des entités numériquement distinctes. Nous avons vu que cette position récupère la *bundle theory* qu'il soutenait initialement, selon laquelle nous sommes essentiellement des personnes dont l'existence n'est que conceptuelle. Elle transforme cependant notre relation à l'animal humain : plutôt que de supposer implicitement que la personne est constituée d'un animal, c'est désormais l'animal qui a une personne comme partie. Ce renversement du constitutionnalisme permettrait à Parfit de se soustraire au problème du dédoublement des entités pensantes qu'entraîne normalement la distinction lockéenne. Seule la personne penserait ; l'animal ne ferait qu'hériter des facultés de celle-ci et ne les exercerait alors qu'indirectement, par l'entremise de cette partie pensante. De plus, ce repositionnement de la personne justifierait l'idée que nous sommes la personne plutôt que l'animal, puisque nous nous concevons comme l'entité qui pense par elle-même plutôt que de manière dérivée. Cette position demeure d'ailleurs compatible avec la *Transplant Intuition*, la vie mentale de la personne étant réalisée dans son cerveau.

Par contre, la position finale de Parfit n'est pas sans ses problèmes. Nous avons soutenu que si l'on attribuait l'accomplissement de chaque activité à une partie de l'animal plutôt qu'au tout, comme le fait Parfit, cela nous empêcherait de comprendre le comportement des êtres vivants comme celui d'agents unifiés, ayant une cohérence interne. Pire, les êtres vivants seraient eux-mêmes réduits à des *bundles* conceptuels de parties d'organisme, suivant le réductionnisme constitutif. À cet égard, nous avons défendu l'idée qu'un sujet conceptuel ou grammatical – personne ou animal – ne pourrait être la source directe d'aucune activité, y compris la pensée, contrairement à ce que soutient Parfit. Seule une entité réelle pourrait produire des effets réels, faisant une

différence dans ce qui a lieu. Nous en avons conclu que le meilleur candidat pour l'entité pensante que nous sommes, c'est l'animal humain conçu comme entité réelle, plutôt que comme entité conceptuelle ou comme partie de l'animal.

Nous avons par conséquent proposé une position qui prend cette conclusion comme point de départ. Nous serions ainsi des animaux, et comme pour tous ceux-ci, notre critère d'identité serait biologique, faisant de cette position un animalisme. Afin d'éviter les absurdités qu'entraîne la distinction lockéenne, nous l'avons rejetée comme les autres animalistes, réduisant la personne humaine à une manière de saisir l'animal rationnel que nous sommes. Cependant, contrairement à l'animalisme classique qui rompt avec l'intuition lockéenne voulant que la conscience soit essentielle à notre nature ontologique, nous avons cherché à la maintenir, à l'instar de Parfit, mais cette fois-ci en la récupérant à même l'approche biologique. C'est pourquoi nous avons nommé cette position l'animalisme *compatibiliste*. Afin qu'une telle réconciliation des contraires soit possible et que nous puissions nous concevoir comme des animaux essentiellement conscients, nous avons soutenu que, plutôt que de parler de la persistance de la conscience en termes de *contenus* mentaux comme le font la majorité des lockéens, il faut nous en tenir aux seules *capacités*. Cette approche permet d'intégrer la conscience au critère d'identité biologique des animaux ainsi doués, et d'accommoder la possibilité d'une perte de contenus mentaux sans que ça ne fasse de nous quelqu'un d'autre. Le compatibilisme requiert donc, comme la deuxième position de Parfit, que le support de la conscience, le cerveau dans le cas présent, soit la partie essentielle de notre organisme. Si les contenus mentaux ne font alors pas partie de notre identité (animale), nous avons avancé qu'ils pourraient très bien être inclus dans l'identité personnelle comme clause additionnelle au critère biologique.

De plus, nous avons suggéré que l'animalisme compatibiliste pourrait rendre compte dans l'ensemble de nos croyances intuitives à propos de notre identité, à savoir qu'elle est déterminée, qu'elle importe, et que son maintien ou non atteste d'une différence réelle et n'est pas question de degrés. Cette position serait ainsi une version valide de la *ego theory*, qui mériterait d'être développée et adéquatement défendue. De toutes les positions que nous avons évaluées dans ce mémoire, elle semble d'ailleurs être la seule capable d'offrir une compréhension unifiée de nous-mêmes, qui sache nous saisir comme

un objet matériel immanent à la nature, et nous attribuer les qualificatifs d'*animal*, de *personne*, de *sujet d'expérience*, d'*agent* et de *penseur* plutôt que d'en accorder certains à une entité séparable de nous-mêmes. La seule manière de rendre tous ces descripteurs compatibles semble être de tous les subsumer sous le concept d'*animal*.

Cela dit, nous avons pu constater dans ce mémoire que le questionnement sur notre nature ontologique et notre identité dépend de questions plus générales à propos de l'ontologie et de la biologie. Nous n'avons pas pu en traiter directement ici, et nous avons donc fait maintes assumptions dans l'argumentation, plausibles certes mais qui méritent d'être évaluées davantage. Notamment, la question de savoir ce qu'est une entité *réelle* a été centrale dans notre développement. Nous nous sommes demandé à divers moments si une substance, un organisme, un organe fonctionnel, du tissu organique, un événement (mental), un objet matériel, et même un *bundle* d'éléments reliés pouvaient être considérés réels ou non. Rattachée à cette interrogation se trouve celle de la *réduction ontologique*. Nous avons supposé, à la suite de Parfit, que ces concepts expriment des idées opposées : une entité réduite ne serait pas réelle et inversement. Leur relation pourrait être conçue autrement par contre : les objets réels pourraient être des *bundles* de particules adéquatement reliées. C'est ce même questionnement qui anime d'ailleurs van Inwagen lorsqu'il se demande quelles sont les conditions pour que des particules élémentaires se *composent* en un objet réel, plutôt que de ne former qu'un *bundle* de particules agencées d'une certaine manière. L'opposition *composition/constitution* que nous avons relevée nécessiterait une investigation plus approfondie.

De plus, nous avons employé à profusion les concepts d'*organisme* et de *vie*, sans trop nous attarder à les préciser. D'ailleurs, le débat moderne se base encore sur les définitions qu'en donne Locke, comme pour la *personne*. Nous avons ainsi pu examiner à la fois des positions qui conçoivent que le corps décérébré est un organisme vivant, et d'autres qui ne le considèrent que comme du tissu vivant, mais qui prennent le cerveau pour un organisme. Nous avons même soutenu l'idée que la conscience serait une faculté essentielle au fonctionnement de l'organisme humain, sans laquelle sa vie prendrait fin. Un critère pour distinguer entre un organisme vivant et un non-organisme qui remplit toutefois certaines fonctions biologiques vitales serait nécessaire. Bref, les affirmations que mettent en jeu les diverses conceptions de l'être humain demandent dans bien des cas

à être évaluées à la lumière de recherches empiriques.

En définitive, il semble qu'il ne soit possible d'avancer dans le débat sur ce que nous sommes et sur nos conditions de persistance qu'en se faisant une meilleure idée de la manière dont la nature, c'est-à-dire le réel, est structuré. Tout comme la critique des animalistes a montré que pour débattre adéquatement du critère de notre identité, il faut avant tout déterminer quelle sorte d'entité nous sommes, la présente investigation démontre que l'ontologie des êtres humains doit être replacée dans le cadre d'une ontologie générale afin d'être cohérente.

Toujours est-il qu'Olson ne semble pas avoir tort en affirmant ceci :

« If we suppose that we have our criterion of identity in common with gods, angels, or electronic thinking machines, rather than with chimpanzees, horses, and oysters, it seems to me that we are pretending to be something that we are not, and refusing to accept our place as biological beings in the natural world. This refusal has a long and fascinating history, and I should not know how to defend my conviction that it is a delusion (and a rather arrogant one at that). » (Olson 1994 : 186)

Si nous sommes des animaux, et que nous ne pouvons donc survivre à notre mort biologique, aussi bien faire le mieux avec la seule vie qui nous est accordée.

ANNEXE A

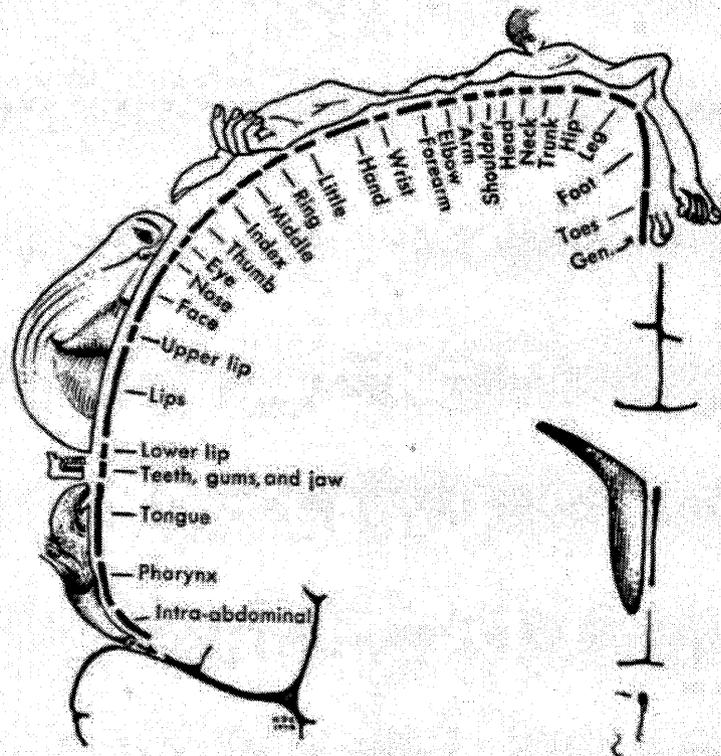


Fig. 17. Sensory homunculus. The right side of the figurine is laid upon a cross section of the hemisphere, drawn somewhat in proportion to the extent of sensory cortex devoted to it. The length of the underlying block lines indicates more accurately the comparative extent of each representation, as in Fig. 9. Compare with Fig. 22.

Homonculus sensoriel

(Penfield & Rasmussen 1950 : 44)

BIBLIOGRAPHIE

- Allen, Colin et Trestman, Michael. (2017). « Animal Consciousness » dans *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (dir.).
[<https://plato.stanford.edu/archives/win2017/entries/consciousness-animal/>]
- Belzer, Marvin. (1996). « Notes on Relation R », dans *Analysis*, 56(1), p. 56-62.
- Blackburn, Simon. (1997). « Has Kant Refuted Parfit ? ». [Chapitre de livre]. Dans Dancy, Jonathan (dir.), *Reading Parfit* (chap. 9, p. 180-199). Oxford : Blackwell.
- Blatti, Stephan. (2012). « A new argument for animalism » dans *Analysis*, 72(4), p. 685-690.
- . (2016). « Animalism » dans *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (dir.).
[<https://plato.stanford.edu/archives/win2016/entries/animalism/>]
- Boly, Melanie, *et al.* (2012). « Consciousness in humans and non-human animals : recent advances and future directions », dans *Frontiers in Psychology*, 4(625), p. 1-20.
- Cabanac, Michel *et al.* (2009). « The emergence of consciousness in phylogeny » dans *Behavioural Brain Research*, 198(2), p. 267-272.
- Campbell, John. (1992). « The First Person : The Reductionist View of the Self ». [Chapitre de livre]. Dans D. Charles & K. Lennon (dir.), *Reduction, Explanation and Realism* (chap. 14, p. 381-419). New York : Oxford University Press (Clarendon Press).
- Cassam, Quassim. (1992). « Reductionism and First-Person Thinking ». [Chapitre de livre]. Dans D. Charles & K. Lennon (dir.), *Reduction, Explanation and Realism* (chap. 13, p. 361-380). New York : Oxford University Press (Clarendon Press).
- Chisholm, Roderick. (1976). *Person and Object* (2^e éd. 1979). La Salle, Illinois : Open Court Publishing Company.
- Dainton, Barry. (1998). Review of *The Human Animal* by Eric Olson dans *Mind*, New Series, 107(427), p. 679-682.
- Dancy, Jonathan (dir.). (1997). *Reading Parfit*. Oxford : Blackwell.
- Descartes, René. (1953). « Lettre à Morus, 5 Février 1649 ». [Chapitre de livre]. Dans G. Lewis (introduction et notes), *Correspondance avec Arnauld et Morus* (p. 110-127). Paris : Vrin.

- Dominus, Susan. (2011, 25 mai). « Could Conjoined Twins Share a Mind? » dans *The New York Times Magazine* (Lauren Kern, dir.).
[\[http://www.nytimes.com/2011/05/29/magazine/could-conjoined-twins-share-a-mind.html\]](http://www.nytimes.com/2011/05/29/magazine/could-conjoined-twins-share-a-mind.html)
- Ehring, Douglas. (1999). « Fission, Fusion and the Parfit Revolution », dans *Philosophical Studies*, 94(3), p. 329-332.
- Evans, Gareth. (1978). « Can There Be Vague Objects ? » dans *Analysis*, 38(4) p. 208.
- Gallagher, Shaun (dir.). (2011). *The Oxford Handbook of the Self*. New York : Oxford University Press.
- Gazzaniga, Michael. (2018). *The Consciousness Instinct*. New York : Farrar, Straus and Giroux.
- Gillett, Carl. (2013). « What you are and the evolution of organs, souls and superorganisms : a reply to Blatti » in *Analysis*, 73(2), p. 271-279.
- Griffin, Donald. (2000). « Scientific Approaches to Animal Consciousness » dans *American Zoologist*, 40(6), p. 889-892.
- Harrison, Peter. (1992). « Descartes on Animals » dans *The Philosophical Quarterly*, 42(167), p. 219-227.
- Hume, David. (2007). *A Treatise of Human Nature*. Dans Norton & Norton, (dir.), *The Clarendon Edition of the Works of David Hume, vol. 1 : Texts* (p. 1-402). New York : Oxford University Press (Clarendon Press).
- James, William. (1884). « What is an Emotion ? » dans *Mind*, 9(34), p. 188-205.
- Johnston, Mark. (1997). « Human Concerns without Superlative Selves ». [Chapitre de livre]. Dans Dancy, Jonathan (dir.), *Reading Parfit* (chap. 8, p. 149-179). Oxford : Blackwell.
- . (2007). « "Human Beings" Revisited : My Body is Not an Animal ». [Chapitre de livre]. Dans D. Zimmerman (dir.), *Oxford Studies in Metaphysics vol. 3* (p. 33-74). New York : Oxford University Press.
- Kim, Jaegwon. (1998). *Mind in a Physical World*. Cambridge : MIT Press.
- Lee, Patrick et Grisez, Germain. (2012). « Total Brain Death : A reply to Alan Shewmon » dans *Bioethics*, 26(5), p. 275-284.
- Leibniz, Gottfried Wilhelm. (1966). *Nouveaux essais sur l'entendement humain*, Paris : Garnier Flammarion.
- Lewis, David. (1976). « Survival and Identity ». [Chapitre de livre]. Dans A. Rorty (dir.), *The Identities of Persons* (chap. 1, p. 17-40). Berkeley, California : University of California Press.
- Locke, John. (1822). *Oeuvres philosophiques de Locke*, tome III (édition revue par J.-F. Thurot). Paris : Firmin Didot.
- . (1975). *An Essay Concerning Human Understanding* (Peter Nidditch, dir., révision 1985). Oxford : Oxford University Press.

- . (2004). *An Essay Concerning Human Understanding*. (édition électronique en anglais moderne par Jonathan Bennett, révision décembre 2012). [www.earlymoderntexts.com]
- Locke, John & Leibniz, Gottfried Wilhelm. (1839). *Oeuvres de Locke et Leibnitz, contenant l'Essai sur l'entendement humain... accompagné de notes* (édition revue par J.-F. Thurot). Paris : Firmin Didot.
- Mackie, David. (1999). « Animalism versus Lockeanism : No Contest » dans *The Philosophical Quarterly*, 49(196), p. 369-376.
- Marlaire, Ruth. (2015). « NASA Ames Reproduces the Building Blocks of Life in Laboratory ». [<https://www.nasa.gov/content/nasa-ames-reproduces-the-building-blocks-of-life-in-laboratory>]
- Maturana, Humberto & Varela, Francisco. (1987). *The Tree of Knowledge* (révision 1998 ; R. Paolucci, trad.). Boston : Shambhala Publications.
- McDowell, John. (1997). « Reductionism and the First Person ». [Chapitre de livre]. Dans Dancy, Jonathan (dir.), *Reading Parfit* (chap. 11, p. 230-250). Oxford : Blackwell.
- McMahan, Jeff. (2002). *The Ethics of Killing : Problems at the Margins of Life*. New York : Oxford University Press.
- Merker, Björn. (2007). « Consciousness Without a Cerebral Cortex: A Challenge for Neuroscience and Medicine » dans *Behavioral and Brain Sciences*, 30(1), p. 63-81.
- Nagel, Thomas. (1971). « Brain bisection and the unity of consciousness ». [Chapitre de livre]. Réimprimé dans T. Nagel, *Mortal Questions* (1979, p. 147-164). Cambridge : Cambridge University Press.
- . (1974). « What is it like to be a bat? ». [Chapitre de livre]. Réimprimé dans T. Nagel, *Mortal Questions* (1979, p. 165-180). Cambridge : Cambridge University Press.
- . (1979). *Mortal Questions*. Cambridge : Cambridge University Press.
- . (1986). *The View From Nowhere*. New York : Oxford University Press.
- . (2012). *Mind & Cosmos*. New York : Oxford University Press.
- Nichols, Peter. (2010). « Substance concepts and personal identity » dans *Philosophical Studies*, 150(2), p. 255-270.
- Noonan, Harold. (1989). *Personal Identity* (2^e éd. 2003). London : Routledge.
- . (2001). « Animalism versus Lockeanism : Reply to Mackie » dans *The Philosophical Quarterly*, 51(202), p. 83-90.
- . (2004). « Are There Vague Objects » dans *Analysis*, 64(2), p. 131-34.
- Olson, Eric. (1994). « Is Psychology Relevant to Personal Identity ? » dans the *Australian Journal of Philosophy*, 72(2), p. 173-186.

- . (1995). « Human People or Human Animals ? » dans *Philosophical Studies*, 80(2), p. 159-181.
- . (1997). *The Human Animal*. New York : Oxford University Press.
- . (2002a). « Thinking Animals and the Reference of 'I' » dans *Philosophical Topics*, 30(1) : *Identity and Individuation*. p. 189-207.
- . (2002b). « What does functionalism tell us about personal identity? » dans *Noûs*, 36(4), p. 682-698.
- . (2003a). « An Argument for Animalism ». [Chapitre de livre]. Dans R. Martin et J. Barresi (dir.), *Personal Identity* (p. 318-334). Oxford : Basil Blackwell.
- . (2003b). « Was Jekyll Hyde ? » dans *Philosophy and Phenomenological Research*, 66(2), p. 328-348.
- . (2006). « Imperfect Identity » dans *Proceedings of the Aristotelian Society*, 106(2), p. 249-266.
- . (2007). *What Are We ? A Study in Personal Ontology*. New York : Oxford University Press.
- . (2015a). « On Parfit's View That We Are Not Human Beings ». [Chapitre de livre]. Dans A. O'Hear (dir.), *The Royal Institute of Philosophy Supplements* vol. 76 : *Mind Self and Person* (p. 39-56). Cambridge University Press.
- . (2015b). « What does it mean to say that we are animals ? » dans the *Journal of Consciousness Studies*, 22(11-12), p. 84-107.
- . (2015c). « Animalism and the Remnant-Person Problem ». [Chapitre de livre]. Dans J. Fonseca & J. Gonçalves (dir.), *Philosophical Perspectives on the Self* (p. 21-40). Bern : Peter Lang.
- . (2016). « The Role of the Brainstem in Personal Identity ». [Chapitre de livre]. Dans A. Blank (dir.), *Animals : New Essays*. Munich : Philosophia.
- . (2017). « Personal Identity » dans *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Edward N. Zalta (dir.).
[<https://plato.stanford.edu/archives/sum2017/entries/identity-personal/>]
- Panksepp, Jaak. (2005). « Affective consciousness : Core emotional feelings in animals and humans », dans *Consciousness and Cognition*, 14(1), p. 30-80.
- . (2010). « Affective consciousness in animals : perspectives on dimensional and primary process » dans *Proceedings : Biological Sciences*, 277(1696), p. 2905-2907.
- . (2011). « The basic emotional circuits of mammalian brains : Do animals have affective lives ? » dans *Neuroscience and Biobehavioral Reviews*, 35(9), p. 1791-1804.
- Parfit, Derek. (1971a). « Personal Identity » in *Philosophical Review*, 80(1), p. 3-27.
- . (1971b). « The Importance of Self-Identity » dans *The Journal of Philosophy*, 68(20), p. 683-690.

- . (1982). « Personal Identity and Rationality » dans *Synthese*, 53(2), p. 227-241.
- . (1984). *Reasons and Persons* (révision 1987). New York : Oxford University Press (Clarendon Press).
- . (1987). « Divided minds and the nature of persons ». [Chapitre de livre]. Dans C. Blakemore et S. Greenfield (dir.), *Mindwaves* (p. 19-26). Oxford : Basil Blackwell.
- . (1995). « The Unimportance of Identity ». [Chapitre de livre]. Dans H. Harris (dir.), *Identity* (chap. 2, p. 13-45). Oxford : Oxford University Press.
- . (1999). « Experiences, Subjects, and Conceptual Schemes » dans *Philosophical Topics*, 26(1-2), p. 217-270.
- . (2007). « Is Personal Identity What Matters ? ». [Manuscrit non-publié]. [http://www.stafforini.com/txt/parfit_-_is_personal_identity_what_matters.pdf]
- . (2008). « Persons, Bodies, and Human Beings ». [Chapitre de livre]. Dans T. Sider, J. Hawthorne, and D. Zimmerman (dir.), *Contemporary Debates in Metaphysics* (p. 177-208). Oxford : Basil Blackwell.
- . (2012). « We Are Not Human Beings » dans *Philosophy*, 87(1), p. 5-28.
- Penfield, Wilder et Rasmussen, Theodore. (1950). *The Cerebral Cortex of Man : A Clinical Study of Localization of Function*. New York : The Macmillan Company.
- Pinto, Yair. (2017a, 26 sep.). « When you split the brain, do you split the person? » dans *Aeon* (site web), P. Weintraub (dir.). [<https://aeon.co/ideas/when-you-split-the-brain-do-you-split-the-person/>]
- Pinto, Yair, *et al.* (2017b). « Split brain: divided perception but undivided consciousness » dans *Brain*, 140(5), p. 1231–1237.
- . (2017c). « The Split-Brain Phenomenon Revisited: A Single Conscious Agent with Split Perception » dans *Trends in Cognitive Sciences*, 21(11), p. 835-851.
- President's Council on Bioethics. (2008). *Controversies in the Determination of Death*. Washington, DC. [<https://bioethicsarchive.georgetown.edu/pcbe/reports/death/>]
- Pyke, Judith (dir.). (2017, 5 nov.). *Inseparable: Ten Years Joined At The Head*. CBC Docs POV. The Canadian Broadcasting Corporation, Inseparable Productions Inc., 44:07 mins. [<http://www.cbc.ca/cbcdocs/pov/episodes/inseparable>]
- Quine, Willard van Orman. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.
- Rettler, Bradley et Bailey, Andrew. (2017). « Object » dans *The Stanford Encyclopedia of Philosophy* (Winter 2017 edition), Edward N. Zalta (dir.). [<https://plato.stanford.edu/archives/win2017/entries/object/>]
- Robinson, Howard. (2014). « Substance » dans *The Stanford Encyclopedia of Philosophy* (Spring 2014 edition), Edward N. Zalta (dir.). [<https://plato.stanford.edu/archives/spr2014/entries/substance/>]

- Russell, Bertrand. (1923). « Vagueness » dans *Australasian Journal of Psychology and Philosophy*, 1(2), p. 84-92.
- Shewmon, Alan. (2001). « The Brain and Somatic Integration: Insights Into the Standard Biological Rationale for Equating “Brain Death” With Death », dans *Journal of Medicine and Philosophy*, 26(5), p. 457-478.
- Shoemaker, Sydney. (1970). « Persons and Their Pasts » dans *American Philosophical Quarterly*, 7(4), p. 269-285.
- . (1984). « Personal Identity : a Materialist Account ». [Chapitre de livre]. Dans S. Shoemaker et R. Swinburne, *Personal Identity* (p. 67-132). Oxford : Basil Blackwell.
- . (1985). « Critical Notice of Parfit's Reasons & Persons » dans *Mind*, New Series, 94(375), p. 443-453.
- . (1986). « Introspection and the Self ». Réimprimé dans S. Shoemaker, *The first person perspective and other essays* (1996, p. 3-24). Cambridge : Cambridge University Press.
- . (1997). « Self and Substance » dans *Philosophical Perspectives*, 11, p. 283-304.
- . (1999a). Review of *The Human Animal* by Eric Olson, dans *Noûs*, 33(3), p. 496-504.
- . (1999b). « Self, Body and Coincidence » dans *Proceedings of the Aristotelian Society, Supplementary Volumes*. Oxford University Press. 73(1), p. 287-332.
- . (2008). « Persons, Animals, and Identity. » dans *Synthese*, 163(3), p. 313-324.
- Snowdon, Paul. (1990). « Persons, Animals, and Ourselves ». [Chapitre de livre]. Dans Christopher Gill (dir.), *The Person and the Human Mind* (p. 83-107). Oxford : Oxford University Press (Clarendon Press).
- Sperry, Roger. (1968). « Hemisphere Deconnection and Unity in Conscious Awareness » dans *American Psychologist*, 23(10), p. 723-733.
- . (1984). « Consciousness, Personal Identity and the Divided Brain » dans *Neuropsychologia*, 22(6), p. 661-673.
- Stoljar, Daniel. (2015). « Physicalism », dans *The Stanford Encyclopedia of Philosophy* (Spring 2015 edition), Edward N. Zalta (dir.).
[<http://plato.stanford.edu/archives/spr2015/entries/physicalism/>]
- Unger, Peter. (1979). « Why There Are No People » dans *Midwest Studies in Philosophy*, 4(1), p. 177-222.
- . (1990). *Identity, Consciousness and Value*. New York : Oxford University Press.
- van Inwagen, Peter. (1990). *Material Beings*. New York : Cornell University Press.
- . (1993). « Précis of *Material Beings* » dans *Philosophy and Phenomenological Research*, 53(3), p. 683-686.

- . (2007). « A Materialist Ontology of the Human Person ». [Chapitre de livre]. Dans van Inwagen & Zimmerman (dir.), *Persons : Human and Divine* (p. 199-215). Oxford : Clarendon Press.
- Vrije Universiteit. (2017, 26 jan.). « Split brain does not lead to split consciousness », dans *Medicalxpress. Neuroscience News*.
[<https://medicalxpress.com/news/2017-01-brain-consciousness.html>]
- Watson, Joshua. (2016). « Thinking Animals and the Thinking Parts Problem » dans *The Philosophical Quarterly*, 66(263), p. 323-340.
- Wiggins, David. (1967). *Identity and Spatio-Temporal Continuity*. Oxford : Basil Blackwell.
- . (2001). *Sameness and Substance Renewed* (2^e éd. 2003). Cambridge : Cambridge University Press.
- . (2005). « Précis of *Sameness and Substance Renewed* » dans *Philosophy and Phenomenological Research*, 71(2), p. 442-448.
- Williams, Bernard. (1956). « Personal Identity and Individuation ». [Chapitre de livre]. Réimprimé dans *Problems of the Self* (1973, p. 1-18). Cambridge : Cambridge University Press.