

UNIVERSITÉ DE NANTES

UFR Lettres et langages

École doctorale « Cognition, Éducation, Interactions »

LE DÉTERMINISME
ET
LA RESPONSABILITÉ
MORALE

THÈSE DE DOCTORAT EN PHILOSOPHIE

soutenue le 24 juin 2011

par François DE MONNERON

Directeur de thèse : Cyrille MICHON

JURY :

Olivier BOULNOIS (EPHE)

Paul CLAVIER (ENS Ulm)

Jean-Pascal ANFRAY (ENS Ulm)

Cyrille MICHON (Université de Nantes)

Sommaire

Introduction	5
I La liberté et la contingence	13
II La responsabilité et le pouvoir des alternatives	59
III L'argument du transfert.....	107
IV L'analyse conditionnelle des pouvoirs et les théories du « moi profond »	175
V Les omissions et le pouvoir des alternatives	215
VI La sensibilité aux raisons et les mondes possibles.....	271
VII Le libre arbitre et le monde physique Déterminisme et indéterminisme.....	323
Conclusion.....	387
Bibliographie	393
Table des matières	403

Introduction

Depuis le XVIII^{ème} siècle, les sciences de la nature ont permis d'expliquer et de prédire les phénomènes avec un succès croissant en énonçant des lois. Ce progrès impressionnant a suscité le désir de réaliser une œuvre comparable dans l'étude des êtres vivants et de l'homme lui-même. Le développement de la psychologie, de la sociologie et des neurosciences ont favorisé l'idée que les comportements humains résultent souvent de causes que nous ne contrôlons pas et dont l'influence obéit à des lois régulières. Par exemple, beaucoup de gens admettent que certains criminels agissent sous l'effet d'un déterminisme psychologique et ne sont pas responsables de leur comportement. Néanmoins, peu de gens iraient jusqu'à dire que la totalité de nos actions résulte d'un déterminisme naturel ou psychologique.

Cette réticence a au moins deux raisons. Premièrement, nous croyons que nous sommes responsables de la plupart de nos actions, ou au moins d'une partie d'entre elles. Or, si nos actions découlaient de facteurs échappant à notre contrôle, nous n'en serions pas responsables. La deuxième raison a été formulée dès l'Antiquité dans ce qu'on appelle « l'argument paresseux » : si les événements futurs sont déjà fixés, l'échec ou le succès de nos entreprises est déjà fixé lui aussi, et il est donc inutile de faire des efforts ou de délibérer en essayant d'atteindre nos buts. Par exemple, si je dois guérir de ma maladie, je guérirai, que j'appelle ou non le médecin. D'une manière plus générale, si nos actions sont entièrement déterminées par des causes, on se demande quelle signification ont les promesses, les conseils, les encouragements, les menaces et les interdictions. Les stoïciens, à cause de leur doctrine nécessitariste, devaient déjà affronter ces difficultés, à la fois du point de vue théorique et du point de vue pratique. Par exemple, Épictète donne ce conseil :

Refuse le serment, si cela se peut, en toute circonstance ; sinon, dans la mesure du possible. (*Manuel*, XXXIII, 5)¹

¹ Trad. Émile Bréhier, p. 1124 dans *Les Stoïciens*, textes traduits par Émile Bréhier, édités sous la direction de Pierre-Marie Schuhl, Bibliothèque de la Pléiade, Paris, 1962.

Ce conseil découle de la croyance au destin et du fait qu'il est normalement imprévisible. Si je promets de faire une certaine action, et que mon destin m'empêche de la réaliser, je viole ma promesse. L'action de promettre devient a priori difficile à justifier dans un monde entièrement gouverné par le destin ou par le déterminisme.

Dès l'Antiquité, la réflexion philosophique s'intéresse aux difficultés engendrées par l'opposition entre la liberté et la nécessité. Au cours de son histoire, elle étudiera trois ou quatre genres de nécessité. Dans l'Antiquité païenne, l'homme croit à l'influence des dieux et des astres sur la destinée des individus. Avec les pères de l'Église apparaît une nouvelle version du problème : en supposant que Dieu est omniscient, comment l'homme peut-il agir librement si toutes ses actions sont connues à l'avance par son Créateur ? Dans la théologie chrétienne, ce problème de la prescience divine s'ajoute à celui du rapport entre la liberté et la grâce : comment la grâce de Dieu peut-elle agir sur l'homme sans lui enlever au moins une partie de son mérite ?

Une troisième version du problème apparaît en même temps que la physique moderne, à l'époque de Galilée et de Newton. Désormais, on se représente la nature comme entièrement régie par des lois que la science se donne pour tâche de connaître. Pour Auguste Comte, cette connaissance des lois permet à la science de remplir sa double fonction d'explication et de prédiction des phénomènes. Au XIX^{ème} et au XX^{ème} siècles, les sciences humaines, mais aussi les sciences de la nature, utilisent les statistiques et le calcul des probabilités. On énonce toujours des lois afin de prédire et d'expliquer, mais faute de mieux, une grande partie des lois sont seulement probabilistes. Cela ne réfute pas la thèse du déterminisme, mais prouve tout de même que la science contemporaine est loin de l'omniscience dont rêvait Laplace.

Du côté des philosophes, « l'antinomie » entre la liberté humaine et les lois de la nature a toujours suscité beaucoup d'intérêt. Pourtant, en ce qui concerne le XX^{ème} siècle, on observe que la littérature sur ce sujet est d'une abondance et d'une cohérence fort inégale selon les temps et les lieux. En France, les réflexions sur ce sujet paraissent ponctuelles et ne font intervenir qu'un petit nombre d'interlocuteurs, quand il ne s'agit pas simplement de recherches solitaires. Le travail le plus marquant est peut-être celui des spiritualistes (Bergson, Ravaisson, Renouvier).

Au contraire, dans la philosophie anglo-saxonne des dernières décennies, un débat nourri et très cohérent s'est mis en place progressivement, faisant intervenir des philosophes de plus en plus nombreux, qui partagent le même vocabulaire et, pour l'essentiel, les mêmes principes de méthode. Ils se divisent en plusieurs courants. À côté des déterministes proprement dits (ou « durs ») se placent les « déterministes doux », selon l'expression de William James². On les appelle plutôt « compatibilistes ». Au sens strict du terme, les compatibilistes ne sont pas toujours déterministes mais soutiennent que le libre arbitre (ou la responsabilité morale) est compatible avec le déterminisme. Ils s'opposent ainsi aux *incompatibilistes* des deux côtés opposés. On distingue en effet deux attitudes opposées parmi eux : pour les incompatibilistes « durs », c'est le libre arbitre qui n'existe pas. Pour les libertariens, au contraire, c'est le déterminisme absolu qui est une fiction.

Quant aux *libertariens*, s'ils sont d'accord sur l'existence du libre arbitre et sur l'inexistence du déterminisme causal, ils ne s'accordent pas, néanmoins, sur la manière dont le libre arbitre s'intègre dans le monde physique et dans les relations de causalité. Le libetarisme de l'agent (*agent-causal libertarianism*) soutient que le libre arbitre est lié à l'existence d'agents qui causent les actions non en vertu d'un état ou d'un événement présent en eux, mais *par eux-mêmes*, en tant que substances. Ils agissent sans y être déterminés par des causes externes ou internes. Cela n'empêche pas de croire que certains états de l'agent, comme ses croyances et ses désirs, l'inclinent vers certains choix ; ils l'inclinent, mais sans le déterminer. Le libetarisme de l'agent a été défendu par Thomas Reid (1877), Roderick Chisholm (1976), Richard Taylor (1958)³, Randolph Clarke (1993, 2003) et Timothy O'Connor (1995b, 2000).

L'autre thèse, le libetarisme des événements (*event-causal libertarianism*), n'admet que des relations causales entre événements. Il soutient que la responsabilité ne suppose pas un agent, mais plutôt des relations causales indéterministes entre des événements. Épicure concevait déjà une idée semblable en associant la liberté à des mouvements d'atomes

² Cf. William James, « The Dilemma of Determinism », *Unitarian Review*, sept. 1884

³ Richard Taylor, "Determination and the Theory of Agency" (1958), in Sidney Hook ed, *Determinism and Freedom in the Age of Modern Physics*, New York University Press, 1958, New York.

dépourvus de cause⁴. Robert Kane (1989, 1996) et Carl Ginet (1990) ont défendu cette idée sous une forme moderne et plus élaborée.

Les libertariens ne sont pas tous incompatibilistes au même degré ni pour les mêmes raisons. Certains, comme Peter Van Inwagen (1983), défendent une version ou une autre du *principe des possibilités alternatives*. Dans sa version la plus simple (notée « PAP »), ce principe dit qu'un agent qui accomplit une action ne peut en être responsable que s'il a eu la possibilité de ne pas l'accomplir. Imaginons un individu jugé pour avoir commis un vol. Si son avocat prouve qu'il ne pouvait pas éviter d'agir ainsi, cet individu ne paraît plus responsable du méfait, et cela quelles que soient les causes qui l'ont empêché d'agir autrement (une drogue, un lavage de cerveau, des menaces de mort, etc.). Cette version du principe a été abondamment critiquée au moyen de contre-exemples, et les libertariens ont dû remplacer PAP par d'autres formulations exprimant la même idée incompatibiliste. Ces libertariens opposent donc le libre arbitre et la nécessité. D'autres libertariens, comme Eleonore Stump (1990) et David Widerker (2006)⁵, sont des « incompatibilistes-source » (*source incompatibilists*). Ils renoncent entièrement au principe des possibilités alternatives, et pensent qu'une action n'a pas besoin d'être contingente pour être libre ; en revanche, elle ne doit pas être causée. Cela suppose de l'indéterminisme à l'origine de chaque action libre.

Reprenons ces deux idées :

1. A est un acte libre (de S) $\rightarrow$ S aurait pu ne pas faire A.
2. A est un acte libre $\rightarrow$ A n'a pas de cause suffisante.

L'incompatibilisme-source rejette 1 en rejetant le principe des possibilités alternatives, et critique le déterminisme en s'appuyant sur 2. Ma position dans ce débat est intermédiaire. Je n'admets pas les arguments de Stump et Widerker contre le principe des possibilités alternatives, mais je préfère m'appuyer sur 2 plutôt que sur 1 pour critiquer le déterminisme. Je pense en effet que la première idée s'explique par la deuxième, et donc s'y réduit. De fait,

⁴ Cf. Épicure, *Lettre à Ménécée*, § 133-135.

⁵ Avant cette date, Widerker défendait la même thèse que Van Inwagen.

dans tous les cas où la nécessité paraît exclure la responsabilité morale⁶, il s'agit d'une nécessité causale. Autrement dit, à chaque fois, l'action est déterminée par des causes. Lorsqu'un événement rend nécessaire un autre événement, il le cause. Donc si un événement extérieur ou antérieur à l'agent rend son action nécessaire, il la cause, et c'est cela, à mon avis, qui exclut la responsabilité morale. Que se passerait-il si l'action était nécessaire en vertu d'une nécessité autre que causale ? S'il s'agissait d'une nécessité logique, l'agent ne serait certainement pas responsable, car personne n'est responsable de faits comme « $2+2 = 4$ » ou « si p implique q , non- q implique non- p ». Mais s'il s'agissait d'une troisième sorte de nécessité, l'agent serait peut-être responsable, car son action ne serait pas causée par une chose ou un événement échappant à son contrôle. Il me semble qu'alors, il en serait responsable.

En raison de cette possibilité, je pense que pour critiquer les thèses déterministes et compatibilistes, il ne faut pas chercher à prouver l'incompatibilité de la nécessité avec la liberté, mais l'incompatibilité du déterminisme causal avec la liberté. J'utiliserai donc une version causale du principe des possibilités alternatives. Ensuite, je défendrai un deuxième principe qui joue un grand rôle dans l'argumentation incompatibiliste : le principe du transfert. Comme le premier, il existe sous plusieurs formes, causales ou non. En voici par exemple une version non-causale et une version causale :

(version non-causale) si p est inévitable, et si p implique q de façon inévitable, alors q est inévitable.

(version causale) si p est inévitable, et si p *cause* q de façon inévitable, alors q est inévitable.

Les différentes versions du principe du transfert servent à fonder l'argument de la conséquence, aussi appelé argument du transfert, ou argument « direct » (certains auteurs le

⁶ Sauf quand il s'agit de nécessité logique ou conceptuelle : par exemple, personne n'est responsable du fait que $2 + 2 = 4$. Autrement, la nécessité qui s'oppose à la liberté de l'agent semble être toujours causale, et le fait qu'une action soit causée par autre chose que l'agent suffit pour qu'il n'en soit pas responsable, même si cette action n'est pas causée de façon nécessaire.

considérant comme plus direct que celui qui part du principe des possibilités alternatives). L'argument du transfert peut se résumer de la manière suivante :

Si le déterminisme est vrai, alors nos actions sont des conséquences des lois de la nature et d'événements appartenant au passé lointain (et même antérieur à notre naissance). Puisque ces lois et ces événements passés échappent à notre contrôle, tout ce qui en est une conséquence échappe aussi à notre contrôle. Donc nos actions échappent à notre contrôle, et nous n'en sommes pas responsables.

Par exemple, supposons qu'il y ait une éclipse de soleil. Je ne suis pas responsable de cette éclipse parce qu'elle découle de lois physiques et de faits passés sur lesquels je n'ai aucun contrôle. Or si le déterminisme est vrai, on peut appliquer le même genre de raisonnements à toutes les actions dont nous nous croyons normalement responsables. Par exemple, imaginons que je vole une voiture. Le déterminisme est la thèse selon laquelle tout événement est une conséquence nécessaire d'événements passés en vertu des lois de la nature. Si le déterminisme est vrai, cette action est la conséquence inévitable, en vertu des lois de la nature, de certains événements mentaux et physiques découlant eux-mêmes d'événements antérieurs, par exemple la formation d'une certaine propriété dans mes gènes. Puisque je ne suis pas responsable de mes propriétés génétiques, ni des lois de la nature, je ne suis pas responsable de ce qui en découle. Donc je ne suis pas responsable de mon méfait.

Comme PAP, les versions les plus simples du principe du transfert ont été abondamment critiquées par les philosophes compatibilistes. Je soutiendrai néanmoins que plusieurs formulations de ce principe échappent à leurs objections. Ainsi, ma thèse sera fondée en partie sur les deux principes les plus communément reçus chez les libertariens : le principe du transfert, et celui des possibilités alternatives. Dans les chapitres suivants (chapitres IV à VI), j'établirai une typologie des théories compatibilistes, et je les critiquerai par des objections indépendantes de ces deux principes.

Dans le premier chapitre, je définirai les notions les plus fondamentales, en tâchant d'expliquer mon hypothèse selon laquelle la nécessité ne s'oppose à la liberté que lorsqu'elle est de nature causale. Dans les deux chapitres suivants, je défendrai deux arguments incompatibilistes : dans le chapitre II, l'argument des possibilités alternatives, et dans le

chapitre III, l'argument du transfert. Dans le chapitre IV, je critiquerai plusieurs théories compatibilistes destinées à montrer que la responsabilité morale peut se comprendre indépendamment du principe des possibilités alternatives ou de l'idée que nous sommes la source ultime de nos actions. Certaines de ces théories s'appuient sur l'analyse conditionnelle des capacités, selon laquelle « A est capable de faire x » signifie « A ferait x s'il le voulait » ; d'autres se fondent sur l'idée qu'on est responsable lorsqu'on agit en accord avec ses désirs profonds (ou son « moi profond »). Dans le chapitre V, j'étudierai les conditions de la responsabilité dans les omissions, pour soutenir qu'aucune théorie compatibiliste ne peut en rendre compte de manière satisfaisante. Lorsqu'un individu est responsable d'un événement par omission, il ne *cause* pas cet événement de la même manière que s'il s'agissait d'une action. Or il est difficile, pour les théories compatibilistes, de définir cette relation causale de manière satisfaisante, sans concéder que l'agent est responsable de l'événement parce qu'il avait la *possibilité* de l'empêcher. Dans les cas d'omissions, la responsabilité morale semble liée à l'existence d'une possibilité alternative. Dans le chapitre VI, je me concentrerai sur un certain type de compatibilisme, qui analyse les capacités de l'agent en termes de *mondes possibles*. Selon cette théorie, un agent est responsable d'une action mauvaise s'il y a des situations possibles proches de la situation actuelle où il ne fait pas cette action. Je soutiendrai que cette solution est incohérente et ne rend pas compte, elle non plus, de tous les cas de responsabilité morale. Elle fait dépendre la responsabilité morale d'une disposition bonne que l'agent devrait posséder à un degré normal. Or il n'y a pas de limite nette entre ce degré normal et la même disposition possédée à un degré plus faible, alors qu'il y en a justement une entre la responsabilité morale et l'absence totale de responsabilité. Enfin, dans le chapitre VII, je montrerai pourquoi le libre arbitre est une capacité mystérieuse. Il est difficile de savoir comment cette faculté s'insère dans le monde matériel et dans la causalité physique. Elle suppose probablement de l'indéterminisme physique à l'intérieur de l'agent, or l'existence d'un processus indéterminé ne favorise pas nécessairement la liberté. Si un tel processus est aléatoire, il risque de limiter le contrôle exercé par l'agent sur ses propres actions.

AVERTISSEMENT

Très peu de textes de philosophie analytique sur le libre arbitre ou sur la responsabilité morale ont été traduits en français. En conséquence, les citations traduites de l'anglais (ou d'une autre langue) seront toujours traduites par moi, sauf indication contraire dans les notes de bas de page ou dans la bibliographie.

I La liberté et la contingence

Dans ce chapitre, j'exposerai d'abord un principe de méthode important pour savoir comment traiter nos intuitions sur la liberté et la responsabilité morale. Ensuite, jusqu'à la section 1.4, je définirai le libre arbitre et d'autres notions de base à partir de telles intuitions. J'exposerai enfin une difficulté liée à une sorte de concurrence entre les deux grandes intuitions libertariennes : on se demande si la responsabilité morale est liée plutôt à la possibilité d'agir autrement, ou à une certaine absence de cause. Pour différentes raisons, que j'exposerai dans les sections 1.6 à 1.11, je choisirai la deuxième hypothèse. Je soutiendrai que la première condition peut se réduire à la deuxième et qu'il est prudent de ne pas supposer un lien conceptuel entre la responsabilité morale et la contingence.

1.1 La liberté et le libre arbitre

Dans le débat entre les compatibilistes et les incompatibilistes, les uns comme les autres font très souvent appel à nos « intuitions ». Certains exemples donnent lieu à des conflits d'interprétation dans lesquels chacun se réfère à nos « jugements spontanés » sur la responsabilité de l'agent. Cela nous entraîne vers deux questions : peut-on éviter les dialogues de sourds, et faut-il nous fier à nos intuitions ?

Je pense qu'il faut répondre « non » à la première question. Il est toujours possible qu'un débat se transforme en dialogue de sourds, car on ne peut pas tout démontrer. Comme le dit Aristote¹, toute démonstration repose ultimement sur des principes premiers indémontrables. Les principes de la logique et des mathématiques sont de cet ordre, mais aussi certains principes moraux. Pour défendre un principe général indémontrable, il est néanmoins possible de présenter des exemples. Varier au maximum les exemples est le moyen indispensable pour

¹ Cf. Aristote, *Métaphysique*, Γ, 4, où Aristote défend le principe de non-contradiction : « il est absolument impossible de tout démontrer : on irait à l'infini, de telle sorte que, même ainsi, il n'y aurait pas de démonstration » (p. 197 du tome 1 dans la traduction de J. Tricot, Vrin, Paris, 1953).

tester un principe général qui n'est pas évident pour tout le monde. Mais les exemples utilisés pour ou contre un tel principe donnent eux-mêmes lieu à des conflits d'intuitions, et il peut arriver un moment où on a épuisé toutes les ressources de l'argumentation sans parvenir à un accord.

Je pense qu'il faut alors appliquer le principe de Reid appelé « principe de crédulité » :

(PC) Il faut tenir pour vrai ce qui paraît tel, sauf s'il y a une évidence contraire et plus forte.

Lorsqu'une affirmation n'est pas clairement évidente, on l'évalue en la comparant avec l'affirmation contraire. Par exemple, si « p » n'est pas clairement évidente, nous devons nous demander quelle est l'affirmation la plus probable entre « p » et « non-p ». C'est alors qu'intervient le principe de crédulité : l'affirmation la plus probable est celle qui a le plus d'évidence pour elle. La nature de cette apparence (ou évidence) varie selon les cas. Le principe de crédulité s'applique d'abord à l'évidence sensible. Par exemple, si je vois un étang, je dois me fier à l'apparence qu'il y a un étang devant moi, sauf si j'ai une bonne raison de penser que mes sens me trompent. S'il fait très chaud, il est possible que ma vue soit trompée par un mirage, mais dans ce cas, mon doute repose lui-même sur l'apparence sensible, car si je sais qu'il fait chaud, c'est par la vue et le toucher.

PC implique donc que nos facultés de connaissances sont normalement fiables, et que nous ne pouvons pas douter raisonnablement des informations qu'elles nous donnent sans nous appuyer sur d'autres informations qui viennent d'elles. Si nous n'imposons pas cette limite au doute, il n'en a aucune ; il détruit nos croyances sans discernement et sans pouvoir fonder une seule connaissance.

Mais l'application de PC aux débats sur la responsabilité morale n'est pas toujours aisée. Par exemple, l'idée selon laquelle on est responsable d'une action seulement si on pouvait ne pas la faire est-elle plausible *a priori* ? Beaucoup de philosophes pensent qu'elle exprime en effet une intuition première sur la responsabilité morale. Si c'est le cas, il faut la tenir pour vraie tant qu'elle n'est pas contredite par un exemple ou un argument d'une vraisemblance plus forte. Mais si quelqu'un y voit une croyance douteuse liée à un contexte culturel particulier,

que peut-on lui répondre ? Lorsque mes yeux reçoivent l'image d'un étang, je sais qu'ils reçoivent bien cette image, que l'image d'un étang se présente donc à mon sens visuel, et qu'une personne placée au même endroit que moi verrait la même chose. Ainsi, je suis certain d'avoir affaire à une intuition : une apparence qui s'impose à moi et que je n'ai pas élaborée par mon activité intellectuelle. Mais lorsqu'un individu *paraît* responsable de son action, il est difficile de savoir si cette apparence correspond à ma première impression (naïve et neutre) ou si elle est déjà influencée par ma théorie sur les conditions de la responsabilité morale. Et d'autres personnes peuvent ne pas « voir » la même chose que moi.

Je pense que lorsqu'il y a un conflit d'intuitions face à un principe général, la meilleure méthode pour celui qui veut défendre ce principe est d'expliquer la force apparente des contre-exemples en montrant qu'elle relève de l'illusion, et ensuite, de critiquer par des contre-exemples tous les principes en concurrence avec ce principe général. Pour défendre les grands principes libertariens, il faut neutraliser les contre-exemples en montrant que, bien interprétés, ils ne s'opposent pas à ces principes. C'est ce que je ferai dans les chapitres II et III. Ensuite, il faut critiquer par des contre-exemples les principes généraux proposés par les compatibilistes pour remplacer ces principes, c'est-à-dire réfuter les principes avancés par les compatibilistes pour énoncer les conditions de la responsabilité morale. C'est ce que j'essaierai de faire dans les chapitres IV à VI.

Heureusement, il y a aussi des idées qui sont incontestablement des « intuitions ». Dans ce premier chapitre, je m'appuierai autant que possible sur ces intuitions pour définir le vocabulaire de base. Tout d'abord, qu'appelons-nous la liberté ? Si nous voulons énumérer les sens de ce mot en commençant par les plus faibles, nous pouvons dire que la liberté est d'abord la simple absence d'obstacle. C'est le cas pour les êtres vivants inférieurs à l'homme, ou même pour des choses inertes. Par exemple, en disant qu'un tel « remonte à l'air libre », nous n'attribuons pas une faculté à l'air, mais voulons dire que l'air, à la surface, ne rencontre pas d'obstacle dans ses mouvements. La liberté des animaux est nettement supérieure : un animal est libre quand il peut faire ce qu'il désire. Cette définition se réfère déjà à des facultés : la faculté de désirer, et implicitement, d'autres facultés comme celle de se déplacer.

Mais le débat qui nous intéresse porte sur les facultés spécifiques à l'homme. Son objet n'est pas la liberté comprise comme un droit, ni la maîtrise de soi ou l'autonomie par rapport aux passions, mais une faculté commune à tous les hommes et considérée comme proprement humaine. On admet communément que cette faculté de l'homme suppose ceci :

- (a) il peut faire ce qu'il veut.
- (b) il peut agir par lui-même.
- (c) il peut agir autrement ou se comporter autrement.

Puisque le libre arbitre est la seule forme de liberté qui m'intéresse ici, à chaque fois que j'utiliserai le mot « liberté », il désignera le libre arbitre. Je supposerai dans tous les développements qui suivront cette définition du libre arbitre :

(LA) Le libre arbitre est *la faculté d'être, au moins à certains égards, l'origine ultime de ce qu'on fait.*

Cette faculté implique à mon avis les trois conditions (a) (b), et (c). Je tâcherai de mettre en évidence ce lien dans les prochains chapitres. Mais la question la plus importante pour moi sera de savoir quelles sont les conditions de la responsabilité morale. Si elle suppose la libre arbitre tel que je l'ai défini, elle suppose alors une autonomie *ultime* incompatible avec le déterminisme. Certains philosophes favorables à la thèse déterministe (comme Fischer) estiment que le libre arbitre au sens propre implique des possibilités alternatives (condition c) et que l'homme n'est donc pas *libre*, mais seulement *responsable*, parce qu'il est doué de raison et agit avec une autonomie suffisante. Pour eux, la responsabilité morale ne suppose donc pas le libre arbitre. Ainsi, il est clair qu'en adoptant la définition LA, je ne tranche pas la question portant sur la compatibilité ou l'incompatibilité du déterminisme avec la responsabilité morale.

Je pense qu'une théorie sur les conditions de la responsabilité morale doit accorder une place importante à (a), (b) et (c) comme conditions de la responsabilité morale, car une théorie qui néglige une des ces conditions est fortement contre-intuitive (en écartant une de ces conditions, elle viole le principe de crédulité). De prime abord, la première condition exige moins que la seconde, et celle-ci, moins que la troisième. En effet, on imagine aisément un

individu faisant ce qu'il veut tout en étant déterminé dans son vouloir, par exemple si son esprit est contrôlé par un savant fou. Bien qu'il soit libre au sens (a), il n'agit pas *par lui-même*, donc il n'est pas libre au sens (b). D'autre part, si un individu agit par lui-même, on peut imaginer pourtant qu'il soit déterminé par sa nature, ou par ses actions et ses choix antérieurs. On imagine alors un *déterminisme interne*, c'est-à-dire un processus causal dans lequel les effets et les causes sont dans l'agent, et où les effets sont entièrement déterminés par les causes. On peut soutenir qu'un tel déterminisme exclut la contingence mais non l'action autonome, le comportement de l'agent n'étant pas déterminé par des causes *extérieures*. En fait, les déterministes peuvent même accorder une place à la troisième idée, selon laquelle la liberté suppose une capacité des alternatives. En effet, la possibilité d'agir autrement peut être définie de manière plus ou moins large, comme je l'expliquerai dans la prochaine section. S'il s'agit d'une simple possibilité logique, elle est parfaitement compatible avec le déterminisme causal.

1.2 La contingence et les possibilités alternatives

Beaucoup de gens pensent que l'homme possède la liberté (au sens d'une faculté) et qu'elle implique une forme de contingence dans les actions. La liberté impliquerait que plusieurs possibilités s'offrent à l'agent dans une même situation donnée. Il aurait donc des *possibilités alternatives*. Par exemple, dans dix minutes, je pourrai quitter cette pièce ou y rester. Beaucoup de gens admettent également que la responsabilité morale suppose la possibilité d'agir autrement. Or cette idée se comprend de plusieurs manières.

Lorsqu'on évoque des *possibilités alternatives*, on évoque des états de choses² ou des événements qui n'ont pas lieu mais sont possibles. C'est ce qu'on appelle des situations contrefactuelles. Or cette notion a plusieurs sens. Soit *S* un individu et *p* une proposition vraie

² J'appellerai « fait » ou « état de chose » ce que dénote une proposition lorsqu'il ne s'agit pas nécessairement d'un événement. L'évènement suppose un changement. Quand une proposition n'exprime aucun changement, elle ne peut dénoter un événement, mais seulement un « fait » ou un « état de chose ».

décrivant un état de choses ou un événement. $\sim p$ désigne alors une situation contrefactuelle.
La phrase

(a) « S pourrait faire que *non-p* »

est équivoque. Certains auteurs définissent le mot « peut » de la manière suivante :

(déf.) « S peut faire que *non-p* » $\leftrightarrow$ « S ferait que *non-p* si S le voulait »

C'est ce qu'on appelle l'analyse *conditionnelle* du verbe « pouvoir ». En admettant cette définition, la phrase (a) ne dit pas si *p* peut être évité ou non. Considérons la proposition suivante:

(b) il est inévitable que S fasse que *p*.

Cette proposition est compatible avec :

(c) S ferait que $\sim p$ si S le voulait.

La proposition (b) implique que S ne peut faire que *non-p*. Or, d'après la définition, (c) implique que S peut le faire. Ainsi, la définition conditionnelle du verbe « pouvoir » implique que (b) et (c) sont incompatibles, ce qui est contre-intuitif. En fait, la contradiction disparaît si les propositions (b) et (c) ne portent pas sur le même genre de possibilités. Nous pouvons d'abord distinguer l'impossibilité *physique* et l'impossibilité *logique* ou conceptuelle. L'impossibilité logique est liée au langage (*logos*), au sens même des mots. Par exemple, il ne peut y avoir de cercle carré, tout simplement parce que cela est exclu par les définitions du cercle et du carré. L'impossibilité physique est liée aux lois de la nature, par exemple je ne peux pas voler, ni aller plus vite que la lumière. S'il n'est pas *logiquement* impossible que S fasse *non-p* - donc si ce n'est pas impossible en vertu du sens des mots ou des règles de la logique et des mathématiques - alors c'est *logiquement* possible. Mais ce peut être *logiquement* possible tout en étant *physiquement* impossible. Dans la thèse déterministe, tout ce qui arrive se produit de façon nécessaire, parce que tout événement est une conséquence inévitable d'événements antérieurs. Cela implique qu'il n'y ait qu'un seul futur possible, mais pas forcément un seul futur *logiquement* possible.

En fait, l'idée que *non-p* constitue une possibilité alternative peut même avoir trois sens différents :

- (d) $\sim p$ est *logiquement* possible (compatible avec les lois logiques)
- (e) $\sim p$ est *physiquement* possible (compatible avec les lois physiques)
- (f) $\sim p$ peut arriver compte tenu des lois physiques et des faits déjà réalisés (c'est-à-dire du *passé actuel*)

Ces trois formules se réfèrent à trois formes de contingence. Le déterminisme n'exclut pas les formules (d) et (e), car il admet que certaines situations contrefactuelles sont logiquement possibles, ou compatibles avec les lois physiques. Par contre, il exclut la formule (f), car il affirme que tout ce qui arrive est une conséquence nécessaire du passé actuel et des lois de la nature. Autrement dit, en vertu des lois de la nature, chaque événement est une conséquence inévitable d'événements passés ; il n'a pas d'alternative. Le déterminisme admet donc que certaines situations contrefactuelles sont compatibles avec les lois logiques ou physiques sans être possibles compte tenu du passé actuel. Si le comportement d'un individu est entièrement déterminé par les lois physiques et le passé actuel, il est contingent en un certain sens (puisque'un autre comportement est *logiquement* possible), mais il est tout de même inévitable. Dans le fameux paragraphe XIII du *Discours de Métaphysique*³, Leibniz a voulu montrer que l'homme a souvent la possibilité d'agir autrement même s'il n'y a qu'un seul futur. Selon Leibniz, Dieu a prévu toutes les actions de Jules César et leurs conséquences à long terme. En créant le monde, il a donc prévu que Jules César deviendrait dictateur et renverserait la liberté des Romains, mais l'*imperator* pouvait ne pas franchir le Rubicon et ne pas devenir dictateur, car ses actions n'étaient pas logiquement nécessaires : elles étaient « certaines », mais non « nécessaires ». Finalement, Leibniz distingue la nécessité absolue et la nécessité hypothétique. Il reconnaît donc qu'en un sens, selon sa théorie, toutes les actions

³ G. W. Leibniz, *Discours de Métaphysique*, XIII, p. 36-37. Dans les paragraphes précédents, Leibniz soutient que Dieu a nécessairement créé le meilleur des mondes possibles, parce qu'il est tout-puissant et parfaitement sage. Or il suppose qu'entre deux possibilités qui se présentent à Dieu, il y en a toujours une qui est meilleure que l'autre (deux possibilités ne sont jamais équivalentes du point de vue du bien). De ces deux thèses, il découle apparemment que Dieu a créé un monde entièrement *déterminé*, car il a créé un monde en prévoyant toutes les situations futures et en les choisissant en fonction de leurs avantages par rapport aux autres situations possibles. Leibniz défend donc un déterminisme *théologique*.

de Jules César étaient nécessaires (elles l'étaient « *ex hypothesi* », c'est-à-dire à partir du moment où Dieu avait décidé de créer Jules César).

Dans les développements qui suivront, je soutiendrai que nous avons souvent des « possibilités alternatives » au sens fort, impliquant les trois conditions (d), (e) et (f). *Un fait non-p est une possibilité alternative au sens strict seulement s'il est compatible à la fois avec les lois logiques, les lois physiques, et le passé actuel.* Mais cette précision n'élimine pas toute ambiguïté. La phrase « S peut faire A » est susceptible des deux interprétations suivantes :

1. S peut faire A à t1.
2. S peut faire A (sans restriction temporelle).

Quand on décrit ce que peut faire une personne, on décrit souvent ce qui lui est possible au moment où l'on parle, mais on ne pense pas toujours à un contexte aussi précis. Par exemple, si l'on présente un skieur de talent en disant qu'il peut réaliser une certaine performance en slalom géant, cela ne veut pas dire qu'il puisse la réaliser sur le champ. Imaginons que la conversation ait lieu pendant l'été, en Australie. Il est possible de faire du ski, mais ailleurs. Le skieur a cette capacité au sens large, mais il n'a pas *l'opportunité* de l'exercer. Dans ce cas, la formule « S peut faire A » signifie qu'il peut le faire quand l'occasion se présente, mais non à l'instant où l'on parle⁴.

Cette ambiguïté est importante en ce qui concerne la responsabilité morale. Dans les exemples que nous discuterons, nous nous demanderons si à un instant t1, S aurait pu faire A compte tenu des circonstances. Par exemple, un alcoolique, au hasard de ses flâneries en ville,

⁴ Sur cette distinction, cf. Anthony Kenny, *Will, Freedom and Power*, 1975, p. 131-136. Kenny expose plusieurs différences entre les capacités générales et les opportunités. Par exemple, une capacité « est quelque chose d'interne à l'agent » alors qu'une opportunité « est quelque chose d'externe » (p.133). Les opportunités sont passagères alors que les capacités sont stables (p. 134). J'ai pris l'exemple d'un skieur ayant une capacité générale sans avoir l'opportunité correspondante. Kenny donne aussi un exemple de la situation inverse (p. 136) : le pape Pie XII ne savait pas parler anglais, mais s'exprimait parfois en anglais lors des audiences où il recevait des anglophones. Avec l'aide d'un élocutionniste, il pouvait parler anglais à un groupe d'Américains, en lisant un discours rédigé par un *monsignore* irlandais, mais en faisant cela, il n'exerçait pas la *capacité générale* de parler anglais, car il ne maîtrisait pas cette langue. On peut donc parler anglais sans avoir la capacité de parler anglais. Un autre exemple est celui d'une personne qui touche une cible par hasard, dans un jeu ou au cours d'une chasse, alors qu'elle ne sait pas tirer.

rencontre un ami qui l'invite à boire, et on se demande s'il est possible qu'il ne boive pas. Il faut alors distinguer deux questions : « Est-il possible qu'il ne boive pas ? » et : « Est-il possible qu'il ne boive pas à *partir du moment où son ami l'a invité à boire* ? ». Certains diront qu'il pouvait s'engager dans une autre rue et ne pas voir son ami, mais qu'il ne pouvait plus éviter de boire à partir du moment où il l'avait rencontré. Imaginons que son ami ait beaucoup insisté pour le faire boire. Avant t1, notre alcoolique avait encore la possibilité de ne pas boire ; il l'a perdue après t1, et puisqu'il ne pouvait pas prévoir cette rencontre, nous pensons qu'il n'est pas entièrement responsable de ce qu'il a fait en cédant lamentablement à son penchant pour l'alcool. Dans d'autres cas, l'agent se met *volontairement* dans une situation où il perd la possibilité d'éviter le mal. On le considère alors comme responsable de ce mal malgré l'absence de possibilité alternative. Sa responsabilité est *dérivée*. Un exemple typique de *responsabilité dérivée* est celui d'une personne qui tue quelqu'un en voiture après avoir bu, et qui avait la possibilité de ne pas boire ou de ne pas prendre le volant.

1.3 Les différentes conceptions du libre arbitre

Ces distinctions vont maintenant nous permettre de mieux définir les principales thèses sur le libre arbitre. Reprenons d'abord les trois définitions formulées au début de ce chapitre. S possède le libre arbitre si :

- (a) il peut vouloir ce qu'il veut (autrement dit, vouloir ses volitions ou volontés).
- (b) Il peut vouloir par lui-même.
- (c) Il peut vouloir autrement (ce qui implique des possibilités alternatives).

Cette distinction est l'application au libre arbitre (à la « volonté libre » - *free will*) des conditions que nous avons énoncées plus haut pour l'*action* libre (cf. 1.1). Nous pouvons supposer en effet que le mot « libre » a fondamentalement le même sens quand nous parlons d'*action* libre et de *volonté* libre. Harry Frankfurt a proposé une théorie dans laquelle la liberté humaine est liée à la capacité d'avoir des volitions de second ordre⁵. Si cette théorie est

⁵ Cf. Harry Frankfurt, « Freedom of the Will and the Concept of a Person », *Journal of Philosophy* 68 (1971).

juste, elle permet de soutenir à la fois que nous sommes entièrement déterminés, que nous sommes libres (au sens (a)), et que nous avons le libre arbitre (au sens (a))⁶. Il y a donc une définition du libre arbitre qui n'est pas immédiatement incompatible avec le déterminisme. On appelle compatibilistes les philosophes qui affirment la compatibilité du déterminisme avec le libre arbitre ou avec la responsabilité morale. Cette position a été défendue par de nombreux auteurs dont Keith Lehrer (1966)⁷, Harry Frankfurt (1969), Daniel Dennett (1984), et Ted Honderich (1988). Contrairement à Frankfurt et Lehrer, Dennett et Honderich n'ont pas exploité particulièrement l'idée que le libre arbitre est lié à la capacité de vouloir ses propres désirs.

On appelle incompatibiliste la thèse opposée, selon laquelle le libre arbitre et la responsabilité morale sont incompatibles avec le déterminisme ou même avec la nécessité en général. Telle est la position de Peter Van Inwagen (1983). Sa conception du libre arbitre repose principalement sur la troisième idée que nous avons formulée : cette faculté supposerait la possibilité d'agir autrement. Une telle idée paraît en effet incompatibiliste, à moins d'interpréter la possibilité de vouloir autrement comme une pure possibilité logique, ou physique (sans tenir compte du passé actuel), ou de telle sorte que ce qui est considéré comme possible soit en même temps impossible d'une manière ou d'une autre. La notion de possibilité alternative au sens strict est la notion incompatibiliste par excellence. Leibniz a défendu une position compatibiliste en admettant que la liberté suppose la possibilité d'agir autrement, mais en définissant celle-ci comme une possibilité logique. Or comme nous l'avons vu, un événement peut être logiquement possible sans l'être physiquement. Plus récemment, David Lewis (1981) et Michael Slote (1982), ont soutenu que la liberté est liée à l'existence de possibilités alternatives, mais sans donner à cette expression son sens le plus étroit. Ils ont pu ainsi défendre une thèse compatibiliste. John Martin Fischer (1994), pour sa part, préfère se définir lui-même comme un *semi-compatibiliste*. Il affirme la compatibilité du déterminisme avec la responsabilité morale, mais non avec le libre arbitre, qui requiert à son avis des possibilités alternatives au sens strict.

⁶ Keith Lehrer a repris cette idée en soutenant que la liberté est la capacité de préférer certains désirs à d'autres. Cf. K. Lehrer, "Freedom and the power of preference".

⁷ Keith Lehrer (ed.), *Freedom and Determinism*, Random House, New York, 1966.

La deuxième idée (b) est intermédiaire. La notion d'autonomie s'oppose *a priori* au déterminisme, sauf si elle est comprise en un sens faible. Robert Kane (1989, 1996), Eleonore Stump (1990), et David Widerker (2006), considèrent l'autonomie ultime comme la principale condition de la responsabilité. Ils s'opposent ainsi aux philosophes compatibilistes, qui ne parlent d'autonomie qu'en un sens moins fort. Mais si l'autonomie ultime s'oppose au déterminisme causal, elle ne s'oppose pas aussi clairement à la nécessité. Tout dépend de quelle nécessité il s'agit.

En fait, on peut constater qu'aucune de ces définitions n'est immédiatement incompatible avec le déterminisme. Pour montrer cette incompatibilité, il faut soutenir de préférence la deuxième ou la troisième définition. Si on choisit la deuxième, il faudra prouver en plus que la possibilité d'agir autrement est à comprendre au sens le plus fort (nous discuterons ce point en examinant le principe des possibilités alternatives). Si on choisit la troisième, il faudra montrer que la responsabilité morale suppose une autonomie très forte : agir par soi-même, au sens strict, ce n'est pas seulement avoir un effet sur les choses ou sur soi-même, par exemple en jouant un rôle causal dans la production de ses actions, mais cela suppose d'être une cause première (incausée) d'une partie de ses actions et de ses choix. Cette idée semble impliquer des possibilités alternatives au sens strict, car si une action n'est pas déterminée par des causes, elle n'est pas une conséquence nécessaire du passé actuel et des lois de la nature, donc elle aurait pu ne pas se produire. Néanmoins, il faut tenir compte aussi des autres formes de nécessité (logique et physique). En effet, pour qu'une action ne soit pas nécessaire, il faut non seulement qu'elle ne soit pas causée, mais qu'elle ne soit pas nécessaire d'une autre manière, comme les choix d'un Dieu leibnizien parfaitement bon, qui choisit nécessairement le meilleur.

Je soutiendrai pour ma part que (b) et (c), comprises au sens le plus fort, sont des conditions nécessaires pour qu'un agent soit responsable d'une partie de ses actions. Je donnerai tout son poids à la notion d'autonomie ou d'« autodétermination » (« *self-causation* »), car si un agent peut agir sur lui-même sans être jamais la cause *première* de ses actions, c'est qu'il est entièrement déterminé par des causes extérieures sur lesquelles il n'exerce aucun contrôle.

1.4 Différents sens du mot « pouvoir »

Nous avons vu que le verbe « pouvoir » a plusieurs sens différents. Tout d'abord, en disant qu'un événement E *peut* se produire, on affirme :

1. E est compatible avec les lois logiques. Ou :
2. E est compatible avec les lois physiques. Ou :
3. E est compatible à la fois avec les lois logiques et le passé actuel.

Lorsque nous parlons de ce que peut faire une personne, l'ambiguïté est plus grande encore. Nous pouvons d'abord appliquer la même distinction. S peut faire A si

- (a) A est compatible avec les lois logiques, ou bien si
- (b) A est compatible avec les lois physiques, ou bien si
- (c) A est compatible à la fois avec les lois physiques et le passé actuel

D'autre part, il faut noter que l'affirmation « S peut faire A » a ces deux sens :

- (d) Il est possible que S fasse A.
- (e) S est capable de faire A.

La possibilité exprimée par la formule (d) peut être comprise selon les trois manières indiquées dans les formules (a), (b) et (c). Mais le mot « pouvoir » peut désigner aussi la *capacité*, qu'il ne faut pas confondre avec la simple *possibilité*. Par exemple, nous disons que tel ou tel événement *peut* se produire, mais jamais qu'il *est capable* de se produire. Dans la *Métaphysique*, Aristote a posé une distinction utile pour définir la notion de capacité, par opposition à la simple possibilité : c'est la distinction entre la puissance active et la puissance passive. Ce qui arrive à un individu peut lui arriver de deux façons : en vertu d'une faculté qu'il possède et peut exercer, ou bien par l'action de causes extérieures sur cet individu. Comme l'écrit Aristote,

une première sorte de puissance [...] c'est la puissance passive, c'est-à-dire, dans l'être passif, le principe du changement qu'il est susceptible de subir par l'action d'un autre être, ou de lui-même en tant qu'autre ;

et une autre sorte, c'est l'état de l'être qui n'est pas susceptible d'être modifié dans le sens du pire, ni détruit, par un autre être ou par lui-même en tant qu'autre, sous l'action d'un principe de changement.⁸

La notion de puissance passive est liée au fait que dans certains cas, c'est en subissant une influence extérieure qu'un être « peut » faire telle ou telle chose. Par exemple, je ramasse un caillou à côté d'un puits. Ce caillou peut atteindre le fond du puits, mais on ne dit pas qu'il en soit *capable*, car s'il l'atteint, ce n'est pas par lui-même. Supposons qu'un individu dorme à côté du puits. À moins qu'il soit somnambule, le fait qu'il puisse atteindre le fond du puits pendant son sommeil n'est pas dû à une capacité qu'il posséderait, mais par exemple, à l'action d'une personne mal intentionnée. Il est donc clair qu'il faut distinguer deux sens du verbe « pouvoir » : il exprime tantôt une *faculté ou capacité*, tantôt une simple *possibilité*. L'énoncé « il peut atteindre le fond du puits » peut donc être paraphrasé de deux façons :

1. « Il est capable d'atteindre le fond du puits »
- ou :
2. « Il est possible qu'il atteigne le fond du puits »

Pendant son sommeil à côté du puits, il *peut* atteindre le fond du puits, au sens où cela est *possible* (c), mais non au sens où il en serait *capable* (b). La possibilité qu'un individu fasse quelque chose n'implique aucune capacité s'il fait cette chose simplement sous l'effet de causes extérieures, ou si ce qu'il « fait », comme ici, n'est pas une action. On pourrait donc énoncer cette condition :

Une personne S a la capacité de faire A si A est une action et s'il est causalement possible que S fasse A.

Il ne suffirait pas que S puisse faire A au sens de la possibilité logique, car la capacité de faire quelque chose suppose une sorte de contrôle causal. Par exemple, si je suis capable d'écrire, c'est parce que je contrôle mes doigts. Cette remarque s'applique également à la capacité d'obtenir un certain résultat. « Faire que p », c'est agir ou se comporter de manière à

⁸ Aristote, *Métaphysique* Θ, 1, 1046 a, trad. Jules Tricot.

produire l'événement ou l'état de chose que désigne la proposition « p ». Donc si je suis capable de faire que p en accomplissant l'action A, cela suppose que j'ai un contrôle causal de mon action, et ensuite que mon action cause p.

Il y a des capacités que nous attribuons sans restriction de lieu, de temps, ou de circonstances. Ce sont les *capacités générales*. Par exemple, un individu qui sait skier possède ainsi la capacité générale de skier. Lorsqu'on lui attribue cette capacité, on ne se demande pas s'il peut skier *hic et nunc*. En plein été et loin des pistes de ski, il la possède encore. Mais il ne peut l'exercer que lorsqu'il en a l'*opportunité*. Par exemple, il en a l'opportunité lorsqu'il a accès à des pistes de ski enneigées et à tout le matériel indispensable pour skier. Nous pouvons appeler « capacité particulière » la somme de ces deux conditions :

(déf.) La capacité particulière de faire A suppose la capacité générale de faire A et l'opportunité de faire A.

Ces conditions ne suffisent pas encore pour montrer ce qu'est une capacité au sens strict. Tout d'abord, la capacité générale et la capacité particulière peuvent être ou non des capacités *régulières*⁹. Une *faculté* est une *capacité générale régulière*. Par exemple, on ne sait pas vraiment tirer à la carabine si on manque la cible dans 90% des cas. On a la faculté de faire A, si on accomplit A avec succès un nombre suffisant de fois. De même, pour avoir vraiment la capacité de toucher la cible dans une circonstance particulière, il faut avoir de bonnes chances de la toucher.

Mais la capacité *stricte* requiert aussi que l'agent sache ce qu'il fait. En effet, je peux avoir la capacité *régulière* de taper « xyz777 » au clavier de mon ordinateur, sans savoir que c'est un code secret donnant accès à tous les ordinateurs du Pentagone. Dans ce cas, je ne suis pas *strictement capable* de taper le code secret des ordinateurs du Pentagone. Et pourtant, dans cet exemple, je suis *strictement capable* de taper « xyz777 » au clavier de mon ordinateur, et cette action revient à taper le code secret du Pentagone. En fait, ces deux actions

⁹ Pour définir la capacité régulière, je m'inspire de Tomis Kapitan, « A Master Argument for Incompatibilism ? ». In Kane (2002b).

ne sont pas identiques puisque c'est seulement dans la première que je tape ce code à *mon insu*. « Taper le code secret du Pentagone » et « taper xyz777 » ne sont donc pas seulement deux descriptions différentes de la même action, mais deux actions différentes. Ainsi, la capacité stricte se définit de la manière suivante :

(déf) S a la capacité stricte de faire l'action A à t si et seulement si :

- (a) Il est possible que S fasse A compte tenu des lois logiques, des lois de la nature et du passé actuel relatif à t.
- (b) Si S veut faire A à t, il a des chances suffisantes d'y arriver.
- (c) S'il fait A, il sait qu'il fait A.

La condition (a) implique que d'un point de vue déterministe, S ne peut avoir la capacité stricte de faire A à t que s'il agit effectivement ainsi. Pour un libertarien, au contraire, nous agissons souvent en ayant la capacité stricte d'agir autrement. Dans ce cas, nous avons un pouvoir des alternatives.

1.5 Le pouvoir de faire l'impossible

En prenant l'exemple du tir à la carabine, puis celui du code secret, j'ai montré que *la possibilité d'accomplir une action n'est pas une condition suffisante pour avoir la capacité de l'accomplir*. Inversement, on peut marquer la distinction entre la capacité et la possibilité d'une action en soutenant que celle-ci n'est pas une *condition nécessaire* de celle-là. Autrement dit, l'impossibilité qu'un agent accomplisse une certaine action n'implique pas toujours une incapacité. C'est là une idée beaucoup plus difficile à défendre. Comme je l'ai annoncé en introduction, je supposerai qu'elle est acceptable et je définirai plusieurs mots importants en tenant compte de cette distinction. Dans les prochaines sections, je ne chercherai pas à la prouver, encore moins à la démontrer, mais je tâcherai de l'exposer le plus clairement possible. Ayant fait cela, je présenterai dans les chapitres suivants une critique du déterminisme fondée sur l'idée que la nécessité *causale* est incompatible avec la responsabilité morale. J'éviterai ainsi de supposer que la nécessité *en général* est incompatible avec la responsabilité morale.

Dans certains cas, l'impossibilité d'une action peut sembler compatible avec le pouvoir de l'accomplir. Par exemple :

1. Je ne peux pas torturer quelqu'un pour une petite somme d'argent¹⁰.

Cela veut dire :

2. Il est impossible que je torture quelqu'un pour une petite somme d'argent.

Or on se demande si cela implique :

3. Je suis incapable de torturer quelqu'un pour une petite somme d'argent.

La deuxième proposition exprime une *impossibilité*, tandis que la troisième exprime une *incapacité*. Dans certains cas, il n'est pas clair que la première implique la seconde. Par exemple, suis-je incapable d'insulter une vieille dame dans la rue ? Il semble que non. Supposons qu'elle ne m'ait fait aucun mal. Il est alors impossible que je l'insulte (moralement impossible). Par exemple, mon caractère est tel que je ne puis insulter une personne sans avoir été offensée par elle. Néanmoins, je suis capable de le faire, car j'en ai le *pouvoir*.

Si Dieu existe, il est tentant de croire qu'il est incapable de mentir. Ce qui fonde cette idée, c'est l'impossibilité qu'il mente.¹¹ Mais nous nous trouvons face à un dilemme. Soit Dieu est incapable de mentir, soit il en est capable¹². Il paraît étrange d'opter pour la

¹⁰ Je reprends ici un exemple de Daniel Dennett dans *Elbow Room : The Varieties of Free Will Worth Wanting*, MIT Press, Cambridge (Massachusetts), 1984, p. 133.

¹¹ À première vue, si je suis capable de faire *x*, c'est que je peux le faire. Pourtant, il est facile de trouver des exemples de substances qui ont le pouvoir de faire *x*, mais qui ne peuvent faire *x*. Un morceau de sucre a la faculté de se dissoudre dans l'eau, or on peut imaginer un morceau de sucre qui parvient sur une planète dépourvue d'humidité, et qui ne peut donc pas se dissoudre dans l'eau. De même, une substance putrescible ne devient pas imputrescible par le simple fait qu'on la mette au sec, dans un endroit où il est impossible qu'elle pourrisse. Elle garde sa disposition à pourrir. On peut opposer cet exemple à celui d'un objet solide (comme une boule de pétanque) qui devient fragile parce qu'il est placé dans un milieu où la température est très basse. S'il fait moins 272°C, la boule de pétanque perd sa solidité. Cette fois, on peut le dire, car sa structure moléculaire a changé par l'effet du refroidissement. cf. Bruno Gnassounou, « Pouvoir et possibilité » (2005). Ces deux derniers exemples se trouvent p. 75.

¹² La notion de capacité de mentir n'est pas un non-sens, contrairement à la notion de capacité d'échouer. Nous l'attribuons à certaines personnes et elle ne paraît pas contradictoire, car il y a du positif dans la capacité de mentir (elle suppose de l'intelligence et de l'habileté). Ainsi, puisqu'il existe un pouvoir de mentir, si l'on admet que Dieu est tout-puissant, il faut lui attribuer ce pouvoir. Je partage sur ce point le sentiment de Thomas Reid : « La perfection morale de Dieu consiste, non à ne pas avoir le pouvoir de faire le mal, autrement, comme l'observe avec raison le Docteur Clarke, il n'y aurait pas plus de raison de le remercier pour sa bonté envers nous que pour son éternité ou son immensité ; mais sa perfection morale consiste en cela que, tout en ayant le pouvoir

deuxième hypothèse, car elle lui attribue une imperfection morale. Mais d'un autre côté, si nous disons qu'il est incapable de mentir, nous lui attribuons une incapacité, c'est-à-dire une impuissance.

Comment éviter ici de se contredire ? Je considère que lorsque nous parlons d'une personne, les mots « capacité » et « pouvoir » sont synonymes. En disant que Dieu est tout-puissant, nous affirmons qu'il possède toutes les puissances, c'est-à-dire toutes les facultés ou capacités. Si on nie que Dieu ait la capacité (donc le pouvoir) de mentir, on déclare apparemment qu'il lui manque une partie des aptitudes ou des facultés dont un individu a besoin lorsqu'il a l'intention de mentir (par exemple, il lui manque l'imagination ou l'intelligence nécessaire pour cela). Cette idée est difficilement acceptable dans le cas de Dieu. Une autre solution consiste à dire qu'il a la capacité de mentir, mais qu'il est impossible qu'il s'en serve.

Nous pouvons considérer que la capacité de mentir équivaut à un ensemble de facultés : l'intelligence conceptuelle, la faculté d'exprimer un contenu propositionnel, etc. À partir de là, on peut soutenir que Dieu a tout ce qui est nécessaire pour mentir ; et il est donc capable de faire tout ce qui permet de mentir : connaître le vrai, dire le faux, etc., mais nous pouvons croire qu'il ne peut pas mentir parce qu'il n'en a nullement besoin, ou parce qu'un tel acte suppose de la méchanceté ou une mauvaise intention. Ainsi, le pouvoir d'accomplir un acte n'impliquerait pas la possibilité qu'il soit accompli. La possibilité d'un acte ne serait pas une condition nécessaire du pouvoir de l'accomplir. Je pense que lorsqu'on dit que Dieu est incapable de mentir, on ne veut pas dire qu'il n'en a pas le pouvoir, car ce serait nier sa toute-puissance. C'est donc qu'on ne considère pas la possibilité que Dieu accomplisse un acte comme une condition nécessaire pour qu'il ait le pouvoir de l'accomplir.

de tout faire, un pouvoir auquel on ne peut résister, il exerce ce pouvoir seulement en faisant ce qui est le plus sage et le meilleur » (*Essays on the Active Powers of the Human Mind*, Essay IV, p.262. cf. aussi p.299). Mais Reid a tendance à conclure que Dieu n'est pas soumis à la nécessité, car « être soumis à la nécessité, c'est n'avoir absolument aucun pouvoir » (p.262). Les philosophes du XVIII^{ème} siècle ont en effet tendance à employer le mot « nécessité » en un sens plus fort que nous, donc moins général. cf. aussi Leibniz, *Discours de Métaphysique*, XIII, et *Conversation sur la Liberté et le Destin*, où Leibniz dit que « tout est certain et déterminé » (p.51), mais « qu'il y a de la contingence dans nos actions, et que dans le fond notre prédétermination est une inclination seulement, et non pas une nécessité » (p. 53).

J'ai dit plus haut que la *possibilité* d'accomplir une action n'est pas nécessairement une *capacité* : une personne incompétente peut réaliser des calculs incorrects, mais on ne peut pas pour autant lui attribuer une capacité de réaliser des calculs incorrects. Inversement, il se pourrait que les capacités impliquent parfois des impossibilités : c'est grâce à ses capacités qu'un bon mathématicien *ne peut pas* se tromper dans un calcul simple. En effet, c'est sa capacité de réaliser des calculs très complexes qui l'en empêche. Il serait à mon avis contre-intuitif de dire que ces capacités entraînent des incapacités *pour la seule raison qu'elles entraînent des impossibilités*. Je préfère donc penser qu'elles tendent à exclure les actions qui leur sont contraires sans retirer à l'agent le pouvoir de les accomplir. Ainsi, il serait capable de les accomplir, mais il ne pourrait pas les accomplir.

De la même façon, si Dieu ne peut pas mentir, c'est à cause de ses facultés ; ou encore, si un individu ne peut pas insulter les passants, ce peut être parce qu'il a la capacité de garder son calme et de rester courtois en toutes circonstances, ou celle de distinguer le bien et le mal. Je suppose ici que *toute capacité est un pouvoir*. Alors, toute incapacité, en tant que telle, est une imperfection.

1.6 Trois espèces de capacités

Nous avons vu d'abord qu'il y a une différence entre « S peut faire A » et « S est capable de faire A ». Cette conclusion, en elle-même, ne signifie pas nécessairement qu'il n'y ait aucune relation d'implication, dans un sens ou dans l'autre, entre ces deux énoncés. On pourrait croire que si la possibilité de A n'implique pas nécessairement une capacité, en revanche, la *capacité* de faire A suppose nécessairement la possibilité de faire A. Je viens pourtant d'exposer un argument destiné à montrer que certaines capacités entraînent des impossibilités sans entraîner des incapacités. Dans la continuité de cet argument, je vais maintenant soutenir qu'il n'y a pas de lien conceptuel clair entre l'impossibilité d'une action et l'incapacité de réaliser cette action.

Mon hypothèse de départ est que ce n'est pas *toute nécessité* qui s'oppose à la liberté, mais seulement la *nécessité causale*. Je ne vois en effet aucun exemple de situation où une

nécessité autre que causale enlève à un agent sa responsabilité morale¹³. Ce serait donc l'impossibilité causale (ou « *empêchement* »¹⁴) qui enlèverait la faculté d'accomplir certaines actions. Il se pourrait qu'il y ait d'autres formes de nécessité, en particulier la nécessité des raisons. Je ne vois aucune objection permettant d'exclure *a priori* qu'une nécessité de ce genre intervienne dans le comportement d'un agent. C'est pourquoi je préfère appuyer ma critique du déterminisme sur une opposition entre nécessité causale et liberté plutôt que sur l'opposition entre la nécessité en général et la liberté.

Mon hypothèse est aussi difficile à établir qu'à réfuter. J'essaierai seulement de montrer qu'elle est envisageable et qu'elle évite certaines objections qui peuvent être dirigées contre l'autre hypothèse libertarienne (celle qui suppose un lien conceptuel entre la capacité des alternatives et la contingence *en général*). Dans cette section, je classerai les capacités en trois catégories pour montrer qu'aucune ne suppose *a priori* une absence de nécessité. Je m'intéresserai aussi au fait que le développement d'une capacité et le développement d'une disposition peut rendre certaines actions impossibles sans que l'on puisse montrer que l'agent a perdu la capacité de les accomplir. Dans la section 1.7, je prolongerai cette remarque en prenant l'exemple d'une situation où l'agent est responsable de ce qu'il fait bien qu'il ne puisse pas agir autrement, et je comparerai cet exemple avec un autre dans lequel l'impossibilité en question est de nature causale, pour montrer que c'est seulement dans ce cas que l'agent ne paraît pas responsable. Ensuite (en 1.8), je montrerai quel est le lien entre la nécessité et le fait d'être causé (et ainsi entre les deux grandes intuitions incompatibilistes). En 1.9, je ferai une remarque sur le rapport entre les causes, les obstacles et les défauts, c'est-à-dire tout ce qui limite les capacités. En 1.10, je montrerai que l'idée selon laquelle c'est la nécessité causale qui exclut la responsabilité morale consolide la thèse libertarienne face à deux sortes de contre-exemples utilisés par les compatibilistes. En 11.1 enfin, je montrerai pourquoi l'idée d'une incompatibilité entre le libre arbitre et la nécessité *en général* suppose un lien conceptuel entre l'indifférence et la liberté.

¹³ En dehors du cas très différent de la nécessité logique ou conceptuelle.

¹⁴ Le mot « empêchement » est équivoque. Tel que je l'emploie ici, le mot « empêchement » ne signifie pas qu'il y a un obstacle. Empêcher A de faire x veut dire « faire que A ne puisse pas faire x ». Même l'absence de quelque chose peut empêcher A de faire x, par exemple l'absence d'un moyen de transport.

Tout d'abord, si certaines capacités requièrent la possibilité causale de leur exercice, aucune espèce de capacité n'exclut de façon évidente la nécessité en général. On peut le montrer en distinguant trois sortes de capacités ou de dispositions.

1. Les êtres qui ne sont pas doués d'intelligence ni de volonté (plantes, animaux, et choses inertes) peuvent posséder des dispositions. Par exemple, le sucre a la faculté de se dissoudre dans l'eau, et le peuplier blanc (*populus alba*) a la faculté de développer ses racines traçantes jusqu'à environ 30 mètres de rayon autour du tronc. De telles dispositions se prêtent assez bien à une définition conditionnelle.¹⁵

- x est soluble dans l'eau $\leftrightarrow$ si x est plongé dans l'eau, il se dissout.
- Le *populus alba* a le pouvoir de développer ses racines jusqu'à 30 m de rayon $\leftrightarrow$ si un *populus alba* est planté dans un sol qui lui convient, et si aucun obstacle ne l'empêche de développer ses racines, il développe ses racines jusqu'à un rayon d'environ 30 m.

Il faut remarquer qu'un sucre est soluble dans l'eau même s'il ne se dissout jamais dans l'eau, et même si pour une raison quelconque, il est *impossible* qu'il soit un jour plongé dans l'eau. Par exemple, imaginons qu'à t_1 , un morceau de sucre tombe de la poche d'un cosmonaute au cours d'une expédition sur une planète lointaine absolument dépourvue

¹⁵ Keith Lehrer a inventé un argument contre l'analyse conditionnelle des dispositions. Il part des trois propositions suivantes :

- (1) Si la condition causale C est satisfaite, A fera x .
- (2) Si la condition C n'est pas satisfaite, A ne peut pas faire x .
- (3) C n'est pas satisfaite.

Ces trois propositions sont compatibles logiquement. Or de (2) et de (3), on infère, par *modus ponens*, que (4) A ne peut pas faire x . Or cette conclusion est compatible avec (1). Cela montre la fausseté de la définition conditionnelle « A a le pouvoir de faire $x \leftrightarrow$ Si la condition C est satisfaite, alors A fait x », car il est possible que A ne puisse pas faire x , même si (1) est vrai. Cf. Keith Lehrer (ed.), *Freedom and Determinism*, New York, 1966, p. 196. Comme l'a montré Elizabeth Anscombe, cette objection ne prouve pas vraiment que la définition conditionnelle des dispositions soit fausse, mais que le verbe « pouvoir » a plusieurs sens. Le sucre peut se dissoudre dans l'eau, au sens où il en a la capacité générale (liée à sa structure physique), même s'il ne peut pas se dissoudre dans l'eau, au sens où il n'en a pas l'opportunité. Cela s'applique également aux actions volontaires. cf. note 11 et Elizabeth Anscombe, « Soft Determinism » (1981), ainsi que Bruno Gnassounou, « Possibilité conditionnelle » (2006). B. Gnassounou observe cependant qu'en répondant ainsi à l'objection de Lehrer, on ne prouve pas que la définition conditionnelle soit satisfaisante. Le mot « peut », lorsqu'il désigne une opportunité, « ne saurait être analysé par un conditionnel » (p. 116).

d'humidité. A partir de t_1 , il est impossible que ce sucre se dissolve dans l'eau, mais nous estimons pourtant que la solubilité dans l'eau fait encore partie de ses propriétés dispositionnelles, et cela en raison de sa structure et de ses propriétés chimiques¹⁶. De même, si un peuplier blanc est planté en un lieu où l'étalement de ses vigoureuses racines est limité par la présence de rochers, de telle sorte qu'il ne peut pas les développer jusqu'à un rayon de 30 m, il n'en a pas moins la propriété d'avoir des racines traçantes jusqu'à une distance de 30 m¹⁷.

On dira donc qu'il a ce pouvoir mais qu'il n'a pas l'occasion de l'exercer (et ne *peut* donc pas l'exercer). C'est la forme conditionnelle de la définition qui permet cela. C'est pourquoi si l'homme possède également des pouvoirs admettant une définition conditionnelle, la même conclusion s'appliquera pour eux : un homme pourra garder ces pouvoirs même dans les cas où il sera impossible qu'il les exerce. La particularité d'une disposition telle que la solubilité du sucre est sa compatibilité avec l'impossibilité *causale* de son exercice.

2. Il y a un deuxième genre de facultés qui peut recevoir une définition conditionnelle. Considérons l'énoncé suivant : « je peux parcourir 50 km à pieds ». Cette capacité est liée à la volonté, c'est pourquoi nous la définirons ainsi :

S est capable de faire 50 km à pieds $\leftrightarrow$ (S veut faire 50 km à pieds $\rightarrow$ S fait 50 km à pieds).

Toutes les capacités de ce genre ont une définition de la forme :

S est capable de faire $A \leftrightarrow (s'il\ veut\ faire\ A, il\ fait\ A)$

Contrairement aux dispositions des plantes et des substances chimiques, cette capacité suppose la *possibilité causale* de son exercice. Imaginons un prisonnier : même s'il veut faire 50 km à pieds, il n'en est pas capable puisque les murs et la porte verrouillée l'en empêchent¹⁸. Bien sûr, si nous parlons de la capacité *générale* de faire 50 km à pied, nous

¹⁶ Cf. note 9.

¹⁷ Ou si l'on préfère, la propriété d'avoir des racines traçantes qui ont tendance à pousser jusqu'à 40 mètres du tronc en surface.

¹⁸ Au mieux, je peux faire comme Albert Speer (conseiller de Hitler) pendant sa longue captivité : il entreprit un voyage imaginaire à travers l'Europe, et dont l'itinéraire était prévu jusqu'au détail. Il marchait en boucles dans

pouvons dire qu'elle appartient encore au prisonnier car s'il n'était pas en prison, il pourrait marcher 50 km. Mais si nous parlons de la capacité de faire cela *hic et nunc*, le prisonnier ne la possède pas. D'après ce que nous avons vu en 1.4, la capacité générale se définirait en ajoutant une condition supplémentaire dans le *definiens* :

S est capable de faire A $\leftrightarrow$ Si S voulait faire A, et s'il en avait l'opportunité, alors il ferait A.

Cette capacité se rapproche des simples dispositions dont il était question précédemment, car elle est compatible avec l'impossibilité causale de son exercice. Au contraire, la capacité de marcher 50 km, s'il s'agit d'une capacité particulière (*hic et nunc*), implique la possibilité causale de son exercice. S n'est pas capable de faire A si quelque chose l'en empêche. Faut-il en conclure qu'une telle capacité s'oppose à toute forme de nécessité ? Je pense que rien ne permet de l'affirmer. En effet, si S ne veut pas faire 50 km à pied, il en a toujours la capacité. Or imaginons qu'il ne *puisse* pas avoir cette intention et que cette impossibilité ne soit pas de nature causale. Cela signifie que rien n'empêche S de marcher 50 km, mais que pour une raison quelconque, il est impossible qu'il veuille le faire.

Il est impossible qu'il fasse cela, mais rien ne l'empêche de le faire. Il en est donc capable. Ici, « rien » a le sens de « rien de réel ». Les propositions et les concepts ne sont pas des êtres réels, même si nous pouvons accepter de les considérer comme des êtres. Nous voyons bien en quoi ils diffèrent des substances et des événements : ceux-ci sont réels, ils peuvent être causes ou effets, alors que les notions et les propositions ne sont jamais causes ni effets. De plus, une notion peut être vague ou générale, alors que les êtres réels ne sont jamais des êtres vagues, ni des universels. En gardant à l'esprit cette différence, je fais l'hypothèse que *si l'impossibilité que S fasse A n'est pas due à quelque chose de réel, S possède la capacité de faire A*¹⁹.

la cour de la prison et comptait ses pas pour savoir, à tout moment, où il se serait trouvé en suivant réellement cet itinéraire.

¹⁹ En dehors du cas très particulier de la nécessité logique ou conceptuelle (par exemple, je ne peux pas faire que $2 + 2 = 5$).

3. Ceci nous amène à considérer un troisième genre de capacité : les *dispositions du caractère* que l'agent a acquises par lui-même. Par exemple, il peut développer une disposition à enchaîner les activités de sa journée à toute vitesse, en se projetant en permanence dans l'avenir. C'est ce qu'on peut appeler une disposition à l'hyperactivité. Une personne ayant acquis cette disposition ne peut quasiment pas passer une demie journée à flâner en ville. De même, les *vertus* sont des dispositions, mais contrairement à celles que nous avons déjà évoquées et que l'on peut trouver chez des êtres dépourvus de raison, elles sont liées à la volonté, comme les capacités du deuxième genre. Leur cas est donc intermédiaire entre les deux autres. En tant que disposition, une vertu *incline* celui qui la possède à accomplir un certain genre d'actions. Plus il l'acquiert, moins les actions opposées deviennent probables. Par exemple, plus il acquiert la patience, moins il est colérique ou susceptible, et plus faible est la probabilité qu'il réponde violemment à une petite offense. Dans le cas d'autres vertus comme le courage, la capacité correspondante paraît plus positive au sens où elle n'est pas la capacité de se retenir, mais de *faire* quelque chose. Dans tous les cas, le développement d'une vertu diminue la disposition contraire appelée « vice ».

1.7 La liberté et les dispositions

Contrairement aux capacités du deuxième genre, les vertus ne peuvent pas être définies de façon conditionnelle à partir de la volonté de l'agent. Il serait réducteur de définir la patience en écrivant :

S est patient $\leftrightarrow$ *Si S veut supporter les offenses et les difficultés, il les supporte.*

Cette définition est insuffisante car elle occulte entièrement le fait que la patience est une disposition. Si S est patient, non seulement il est capable de supporter la plupart des difficultés qu'il veut supporter, mais il a aussi tendance à les supporter, donc il est peu probable qu'il ne les supporte pas. Nous pouvons ajouter qu'il a tendance à *vouloir les supporter*. La vertu est en effet une disposition de la volonté ; elle incline la volonté dans un certain sens. Ainsi, l'homme juste a tendance à vouloir la justice, et l'homme tempérant, à refuser les plaisirs excessifs.

Prenons un autre exemple :

[Philalèthe] Philalèthe est un jeune homme sincère. Il a acquis par lui-même cette disposition. S'il se met à décrire une querelle familiale, il y a 95 % de chances²⁰ pour qu'il relate les faits sans les déformer (par commodité, nous ferons comme si une telle probabilité était quantifiable). En effet, dans 95% des situations où on a vu Philalèthe raconter une querelle familiale, il disait la vérité. On interroge Philalèthe, et comme prévu, il dit la vérité.

A-t-il du mérite pour cela, ou sa vertu le rendait-elle quasiment incapable de dire autre chose ? Assurément, la faible probabilité qu'il mente ne diminue pas son mérite. *La responsabilité de l'agent ne semble donc pas proportionnée à la possibilité d'agir autrement (exprimée ici sous forme de probabilités).* Comparons cela avec un autre exemple :

[Superstition] Au cours d'un voyage ethnographique au Brésil, vous discutez avec un indien d'Amazonie. Il est convaincu que les caïmans sont des dieux. Vu son éducation, son manque de formation intellectuelle, et les croyances de son entourage, il y avait 95 % de chances pour qu'il intériorise cette idée bizarre.

Nous avons là deux dispositions acquises, et de force égale. Pourtant, elles ne sont pas de même nature : dans le premier cas, nous estimons que l'agent a du mérite et que son mérite n'est pas diminué par la force de sa disposition. Au contraire, dans le second cas, la force de la disposition est une circonstance atténuante qui nous pousse à excuser l'indien. Il est à peine responsable de sa croyance. Ainsi, dans le premier cas, l'agent est responsable, et dans le second, il ne l'est pas.

La différence tient à la nature des probabilités. Dans le premier cas, elles sont liées à la force des raisons, et non à l'influence de quelque chose ou de quelqu'un. De même qu'il y a une différence entre l'impossibilité causale et l'impossibilité tout court, il y a une différence entre l'improbabilité causale et l'improbabilité tout court. Seule la première diminue la responsabilité de l'agent. Lorsque c'est une vertu acquise par l'agent qui rend l'action opposée

²⁰ Il s'agit de probabilités objectives et non de probabilités épistémiques.

quasiment impossible, on ne peut pas dire que l'agent soit incapable de l'accomplir ; car on le suppose entièrement responsable de ce qu'il fait, et digne d'éloge. Pourtant, il demeure vrai qu'il ne peut quasiment pas agir autrement. Nous voyons ainsi que la capacité de faire une action n'est pas proportionnée à la probabilité de cette action. L'agent peut conserver la capacité de mentir au moment même où il devient quasiment impossible qu'il mente. Ceci suggère que la *capacité* de faire A ne suppose pas nécessairement la *possibilité* que l'agent fasse A.

Il y a cependant une autre façon d'interpréter la différence entre ces deux exemples. On peut soutenir que la responsabilité de Philalèthe est *dérivée* : c'est parce qu'il a eu auparavant la possibilité de ne pas devenir sincère qu'il est responsable de sa propre sincérité et des actions découlant de cette disposition. Ainsi, la possibilité d'agir autrement est bien une condition nécessaire de la liberté et de la responsabilité morale, mais il arrive que les possibilités alternatives ouvertes à l'agent se situent à un moment plus ou moins éloigné dans le passé.

Je pense que pour répondre à cette objection, il faut distinguer le cas présent d'un autre où la responsabilité est clairement dérivée : lorsque l'agent est empêché d'agir autrement.

[drogue] Imaginons qu'un soldat accepte de se droguer avant d'aller combattre. Il connaît les effets de cette drogue. Il sait qu'il aura plus de courage, moins de lucidité et moins de sens critique. Il sait également qu'il pourra devenir féroce et cruel. Une fois drogué, ce soldat commet des atrocités en attaquant un village. Il tue des civils et brûle leurs maisons.

Ici, nous pouvons soutenir que l'agent est responsable de ce qu'il fait même s'il a perdu le contrôle de lui-même et ne peut guère agir autrement. En effet, sa férocité vient du fait qu'il a consommé de la drogue, en toute liberté et en connaissant les effets de la drogue. Nous pourrions imaginer des exemples où il accepte de subir un lavage de cerveau. Le résultat serait le même : étant empêché d'agir autrement, il n'aurait qu'une responsabilité dérivée. Or le

cas de Philalèthe est différent car rien ne l'empêche d'agir autrement²¹ ; nous pouvons donc supposer que Philalèthe a le pouvoir d'agir autrement. En somme, le raisonnement est le suivant :

(1) Si S n'est pas empêché de faire A (par un défaut, une cause ou un obstacle), il a le pouvoir et la capacité de faire A.

(2) Or rien n'empêche Philalèthe de faire A.

(3) Donc il a le pouvoir et la capacité de faire A.

Pourtant, il ne peut pas faire A.

Ce raisonnement me paraît solide. Dans les cas de responsabilité dérivée, l'agent est empêché d'agir autrement par une cause, un défaut ou un obstacle : par exemple, il décide de se droguer, puis accomplit un méfait sous l'influence de la drogue. Ou bien il choisit de ne pas lire une information importante, puis commet une faute grave qu'il aurait pu éviter s'il avait lu cette information. Dans ce genre de cas, on estime que l'agent n'avait pas les moyens ou la capacité d'agir autrement. Or dans le cas des actions vertueuses, il serait étrange de considérer les vertus acquises comme des obstacles ou des défauts qui enlèvent à l'agent une partie de ses moyens d'agir. Pourtant, nous sommes tentés aussi par le raisonnement contraire :

(1) S'il est impossible que S fasse A (ou si la probabilité qu'il fasse A est négligeable), S n'a pas le pouvoir (ou la capacité) de faire A.

(2) Or il est impossible que Philalèthe fasse A.

(3) Donc Philalèthe n'a pas le pouvoir (ou la capacité) de faire A.

²¹ Le fait qu'il ne puisse pas agir autrement n'est pas dû à une cause, un obstacle ou un défaut. S'il l'était, il faudrait considérer qu'il n'y a pas de différence entre les deux exemples. Dans chaque cas, en effet, l'agent aurait une responsabilité seulement dérivée puisqu'il serait empêché d'agir autrement. Mais on pourrait toujours penser que sa responsabilité est liée au fait que son action n'est pas causée (plutôt qu'à une absence de nécessité). En outre, il me paraît étrange de supposer que l'acquisition d'une vertu s'accompagne de la perte d'un pouvoir, c'est-à-dire de l'apparition d'une cause, d'un obstacle ou d'un défaut qui empêche l'agent d'agir autrement.

Dans ce raisonnement, tout me paraît solide, sauf la première proposition, que je ne trouve pas évidente, ou moins évidente que la prémisse (1) du raisonnement précédent. Je préfère donc raisonner de la première façon.

Imaginons que Philalèthe ait la capacité de mentir. C'est une capacité du deuxième genre :

S est capable de mentir $\leftrightarrow$ Si S veut mentir, il ment.

Philalèthe pourrait mentir s'il le voulait. Il a assez d'imagination pour inventer une histoire fausse et, en même temps, vraisemblable. Aucune inhibition, aucune cause interne ne l'empêche de mentir. Mais puisqu'il a acquis la vertu de sincérité, il est devenu quasiment impossible qu'il dise le faux. Apparemment, cette impossibilité n'est pas liée à la perte de sa faculté de mentir, mais plutôt au fait que l'exercice de cette faculté est contrôlé par une vertu. Une faculté du troisième genre (celle d'être sincère) s'est ajoutée à une faculté du deuxième genre (celle de mentir).

1.8 « Causer » et « rendre nécessaire »

Je suppose que ce n'est pas *toute nécessité* qui s'oppose à la liberté, mais seulement la nécessité *causale*. Quelle est donc la différence entre ces deux nécessités ? Nous la comprenons assez bien de façon intuitive, mais pour compléter les explications précédentes et aborder, dans les chapitres suivants, les difficultés liées à la notion de responsabilité morale, je dois formuler un critère de la nécessité causale. Tout d'abord, j'admettrai le principe suivant, qui n'est pas une définition :

(NC – Nécessité causale) Lorsqu'une substance ou un événement rend nécessaire, dans certaines conditions, une autre substance ou un autre événement, elle (ou il) la cause.

Si l'on note ainsi les prédicats : N : rendre nécessaire ; C : causer ; S : substance ; E : événement, alors le principe s'écrit :

Pour tout x , $Sx \rightarrow (Nxy \rightarrow Cxy)$ et : $Ex \rightarrow (Nxy \rightarrow Cxy)$.

Notons bien que ce principe n'exclut pas l'existence de causes non-nécessitantes. Il affirme que tout événement qui rend nécessaire un autre événement le cause, mais il ne dit pas que tout événement qui en cause un autre le rend nécessaire (il ne dit pas que $Sx \rightarrow (Cxy \rightarrow Nxy)$, ni que $Ex \rightarrow (Cxy \rightarrow Nxy)$). Il laisse donc ouverte la question de savoir s'il y a des relations causales indéterministes. Ensuite, pour bien comprendre NC, il est important de voir qu'il y a une différence entre « impliquer » et « rendre nécessaire ». Imaginons un vieux loup de mer capable de prédire le temps²². À chaque fois qu'il annonce une forte houle ou une tempête, sa prédiction s'avère juste. Supposons qu'il soit infaillible dans ses prédictions météorologiques. On peut alors écrire : « Le vieux marin dit qu'il y aura une tempête dans quelques heures » implique « il y aura effectivement une tempête dans quelques heures ». Le premier de ces deux événements implique le deuxième, mais ne le rend pas nécessaire. Si cet événement est nécessaire, c'est sous l'effet de causes climatiques et non sous l'effet de la prédiction. Il y a donc une différence entre impliquer et rendre nécessaire. En tout cas, *rendre nécessaire n'a pas le même sens lorsqu'il s'agit de propositions et lorsqu'il s'agit de substances ou d'événements (voire d'états de choses). Dans le premier cas, nous avons affaire à une relation logique, et dans le second, à une relation causale*²³.

Cette différence apparaît dans les formules suivantes :

- (a) Dieu rend nécessaire y.
- (b) La proposition « Dieu veut y », rend nécessaire la proposition « y ».
- (c) L'état de chose « Dieu veut y » rend nécessaire l'événement y.

A mon avis, les formules (a) et (c) impliquent que y est *causalement* déterminé. Au contraire, (b) ne signifie pas nécessairement que y soit prédéterminé par Dieu.

²² J'emprunte cet exemple à Fischer. Cf. John Martin Fischer, *The Metaphysics of Free Will*, Blackwell, Oxford, 1994, p. 80-81.

²³ En revanche, si la proposition « le marin dit qu'il y aura une tempête dans quelques heures » implique « il y aura effectivement une tempête dans quelques heures », c'est sans doute parce qu'il y a une relation causale déterministe entre certains faits constatés par le marin et la tempête qu'il prédit. Dans ce cas, la relation d'implication entre les deux propositions suppose une relation causale. Pourtant, elle n'exprime pas une relation causale entre les deux faits en question.

De même, on peut dire :

d) « Le vieux marin annonce une tempête » rend nécessaire « il va y avoir une tempête »

mais on ne peut en conclure :

(e) En annonçant la tempête, le vieux marin rend nécessaire la tempête.

Ni :

(f) L'événement dénoté par la proposition « le vieux marin annonce une tempête à t1 » rend nécessaire l'événement dénoté par la proposition « une tempête se produit à t2 ».

En effet, (d) exprime seulement une implication, alors que (e) et (f) expriment une nécessité causale.

Ainsi, lorsqu'un événement rend nécessaire y, cela signifie qu'il le cause. Cependant, en admettant ce principe, je suppose que les événements sont des *particuliers*. Si les événements sont des universels susceptibles de se réaliser de plusieurs manière différentes, il est possible de rendre un événement nécessaire sans le causer, et il faut distinguer ces deux sortes d'influence : déterminer causalement, et causer de manière déterministe. Cette distinction jouera un rôle important dans le débat sur les exemples à la Frankfurt, que j'aborderai au chapitre II. En voici un, que j'emprunte à Van Inwagen²⁴ :

- Gunnar est contrôlé par un neurochirurgien qui a installé un appareil dans son cerveau pour influencer son comportement. Si à t1, Gunnar ne manifeste aucun désir de tirer sur Ridley à t2, l'appareil produit en Gunnar le désir de tirer sur Ridley à t2.
- A t1, Gunnar sort son pistolet de sa poche pour tirer sur Ridley. A t2, il tire sur Ridley, qui meurt aussitôt après. Puisque Gunnar a fait cela de lui-même, l'appareil n'est pas intervenu.

²⁴ Peter Van Inwagen, *An Essay on Free Will*, Oxford University Press, Oxford, 1983, p. 162.

Soit E l'événement « Gunnar tue Ridley à t_2 ». Nous pouvons dire que le neurochirurgien *rend nécessaire* l'événement E à t_2 , et donc le *détermine*. Puisque les moyens qu'il emploie pour garantir ce résultat sont d'ordre causal, il le détermine *causalement*. Cependant, il ne *cause* pas l'événement E, car c'est Gunnar lui-même qui a tué Ridley. Il est donc possible de déterminer causalement un résultat²⁵ sans le causer soi-même. Cependant, cela suppose que deux événements différents dans le détail ou par leur origine causale puisse être considérés comme identiques à partir du moment où ils tombent sous la même description générale. Par la suite, je supposerai que les événements sont toujours des particuliers, c'est la raison pour laquelle ma formulation du principe NC ne tient pas compte des exemples à la Frankfurt. Laissons maintenant ces cas très particulier, et considérons les *causes* de nos actions et de nos croyances.

Revenons à l'exemple « caïmans », où un indien d'Amazonie prend les caïmans pour des dieux. Son éducation est telle qu'il ne peut pas ne pas avoir cette croyance. Puisque cela est dû à ce que des personnes lui ont dit et répété depuis son enfance, nous pouvons penser qu'il est déterminé causalement à recevoir cette croyance. En effet, ce sont des substances et des événements qui l'ont rendue nécessaire. « Rendre nécessaire », ici, veut même dire « causer de manière déterministe ». Cependant, on peut m'objecter que cette croyance n'est pas nécessaire, mais seulement très probable. Je réponds alors que la nature du problème n'en est pas changée : cette remarque nous oblige seulement à introduire des degrés de probabilité dans notre théorie, et à remplacer l'impossibilité causale par l'improbabilité causale. Nous dirons donc ceci :

Plus l'influence d'une substance ou d'un événement x empêche y de faire A , moins il est capable de faire A , et moins il est responsable de ne pas faire A .

²⁵ Nous verrons plus loin (p. 67) que ce résultat est un fait *général*, et non l'événement *particulier* « Ridley meurt » tel qu'il s'est produit dans la situation actuelle. En effet, la présence de l'appareil dans le cerveau de Gunnar garantit le fait que Ridley meure à t_2 , mais non le fait qu'il meure exactement de la même manière que dans la situation actuelle, où Gunnar le tue sans intervention de l'appareil. Pour déterminer causalement un événement particulier, il faut le causer, car un dispositif de type frankfurtien comme celui du neurochirurgien dans l'exemple donné ici ne peut déterminer causalement qu'un fait *général*.

Dans cette formule « empêche y de faire A » peut être remplacé par « réduit la probabilité que y fasse A ». La probabilité dont il s'agit est une probabilité objective évaluée à partir de situations semblables. Par exemple, on observe Philalèthe dans des situations où il peut avoir intérêt à mentir pour se tirer d'affaire, et on constate que dans 95% de ces situations, il ne ment pas. Ici, la probabilité que Philalèthe ne mente pas n'est pas due à l'influence d'une cause. On peut même imaginer qu'il résiste au contraire à des pressions extérieures. A l'inverse, imaginons un individu corrompu qui cède dans 95% des situations où on lui propose de l'argent en échange d'un service malhonnête mais facile à rendre. Bien qu'il subisse une pression extérieure, on le considère comme responsable car c'est *par lui-même* qu'il a développé cette propension à accepter de l'argent pour des services malhonnêtes. Considérons enfin un troisième cas, celui d'une personne qu'on a forcée à se droguer et qui est devenue dépendante. On observe que malgré son désir de vaincre la dépendance, dans 95% des situations où elle voit de la drogue à sa disposition, elle ne résiste pas à la tentation. Ici, la forte probabilité qu'elle cède à la tentation de la drogue est due à l'influence d'une cause. Pour cette raison, elle ne peut être considérée comme pleinement responsable. Je pense donc que ces deux affirmations sont plausibles :

(TNC - thèse non compatibiliste) Plus une chose ou un événement rend l'action alternative improbable, moins l'agent est responsable de ce qu'il fait.

(TNI - thèse non incompatibiliste) Il arrive que l'action alternative soit très improbable et que l'agent soit pleinement responsable de ce qu'il fait, parce que rien ne l'empêche d'agir autrement²⁶.

²⁶ Cf. l'exemple de Philalèthe, p. 40.

1.9 Causes et défauts

Ce qui s'oppose à la responsabilité morale, ce n'est pas toujours l'influence d'une cause extérieure, mais ce peut être aussi la présence d'un *obstacle*. Si une personne voulant sauver un vieillard d'un incendie se heurte à un mur, à une porte infranchissable, ou encore à une barrière de flammes qu'elle ne parvient pas à éteindre, on ne peut certainement pas lui reprocher de ne pas sauver la victime. La présence d'un obstacle insurmontable ôte entièrement la responsabilité morale dans un cas d'omission comme celui-ci. Mais il faut remarquer que l'*absence d'un moyen ou d'une faculté* peut également s'opposer à la responsabilité morale. La non-responsabilité de l'agent peut être due à un défaut, ou pour le dire de façon plus générale, à quelque chose qui *fait défaut*, car ce peut être un outil, un moyen de transport, une information utile, ou n'importe quelle sorte de moyen. Par commodité, j'appellerai « défaut » non seulement les facultés qui manquent à l'agent pour réaliser une action, mais aussi les moyens et instruments dont il ne dispose pas. J'appellerai les premiers « défauts internes », parce qu'il sont internes à l'agent, et les seconds, « défauts externes ».

Parfois, donc, la responsabilité est enlevée ou diminuée par la *présence* de quelque chose (une cause ou un obstacle), et parfois, c'est au contraire l'*absence* de quelque chose qui diminue la responsabilité. En réalité, la distinction entre ces deux cas n'est pas nette. Un défaut n'étant pas une chose, il ne peut guère contraindre l'agent ou lui faire obstacle. Est-ce donc le défaut lui-même qui enlève la responsabilité, ou une cause à laquelle il n'arrive pas à s'opposer parce qu'il lui manque des capacités ou des moyens nécessaires pour cela ? Un obstacle ne limite l'agent que parce qu'il lui manque des facultés. Par exemple, un mur ne serait pas un obstacle pour un individu capable de l'escalader ou de le franchir d'un bond. C'est donc seulement pour celui qui a un défaut (c'est-à-dire qui n'a pas toutes les capacités utiles) qu'il y a des obstacles. De même, si je veux sauver un vieillard d'un incendie mais que je me heurte de tous côtés à des obstacles, c'est bien l'obstacle qui limite ma liberté, mais parce que je n'ai pas les instruments nécessaires pour enfoncer les portes ou éteindre les flammes. Il est probable que les défauts et les obstacles aillent toujours de pair, mais selon les cas, notre attention se dirige vers l'un ou l'autre aspect : l'obstacle, ou le défaut. Dans l'exemple du mur, si nous ne voyons pas qu'il y a un défaut dans les capacités de l'agent, c'est

sans doute parce que pour le voir, il faut imaginer quelque chose d'extravagant (la faculté de bondir à plusieurs mètres de hauteur).

D'après cela, on peut définir de la manière suivante les capacités :

(déf.) S possède la *capacité* d'accomplir l'action A si et seulement si aucun défaut externe ou interne, ni aucune cause externe ou interne, ne l'empêche d'accomplir cette action.

La capacité définie ici est la capacité causale générale ou particulière, selon la terminologie de la section 1.4. La capacité régulière et la capacité stricte requièrent des conditions supplémentaires²⁷. Selon la thèse incompatibiliste, la responsabilité morale suppose qu'en certaines circonstances, l'agent a la capacité d'agir autrement. Cela implique non seulement qu'il puisse agir autrement (au sens de la possibilité) mais que rien ne l'empêche d'agir autrement. Dans ce cas, nous pouvons dire qu'une possibilité alternative lui est *accessible*.

(déf.) En supposant que S ne fait pas *x* mais *y* dans le monde actuel, *x* est une possibilité alternative *accessible* à S (à l'instant *t*) si et seulement si S a la capacité stricte de faire *x* à *t*.

Ayant ainsi défini les concepts et les distinctions de base les plus utiles pour réfléchir sur le déterminisme et le libre arbitre, nous pouvons maintenant entrer dans le vif des débats. Dans les sections 1.5 à 1.9, j'ai expliqué mon hypothèse selon laquelle c'est la nécessité causale (et elle seule) qui s'oppose à la responsabilité morale. Dans les deux prochaines sections, j'exposerai deux raisons qui motivent mon choix. La première est que cette hypothèse minimale est suffisante du point de vue libertarien. Pour critiquer la thèse déterministe, il n'est pas nécessaire de soutenir une hypothèse plus forte. La deuxième raison est que si la responsabilité s'opposait à la nécessité *en général*, il y aurait un lien conceptuel entre la responsabilité et l'indifférence. Or l'indifférence suppose des facteurs qui limitent la responsabilité morale.

²⁷ Cf. *supra*, p. 24.

1.10 La nécessité compatible avec le libre arbitre ?

On distingue parfois deux grandes intuitions libertariennes : premièrement, l'intuition selon laquelle la nécessité s'oppose à la liberté, deuxièmement, l'intuition selon laquelle c'est plutôt le fait d'être causé qui s'oppose à la liberté. Qu'est-ce qui définit le caractère non libre d'une action ? Le fait qu'elle soit nécessaire ou le fait qu'elle soit causée ? De nombreux philosophes ont répondu de la première façon. Van Inwagen oppose la liberté à la nécessité, et soutient que la responsabilité morale suppose la possibilité d'agir autrement. Pour ma part, je pense que c'est le fait d'être causé qui s'oppose à la liberté de l'agent. Je soutiens que la nécessité ne s'oppose pas à la liberté, mais je ne suis pas compatibiliste pour autant, puisque je soutiens que le déterminisme tel que nous le comprenons habituellement (c'est-à-dire comme une sorte de nécessité causale) exclut à la fois la liberté et la responsabilité morale.

Comme j'ai essayé de le montrer, il y a un lien entre ces deux intuitions, de telle sorte qu'il est possible d'en expliquer une à partir de l'autre. J'ai essayé de les unifier dans le principe NC, qu'il n'est pas inutile de rappeler ici :

(NC – Nécessité causale) Lorsqu'une substance ou un événement rend nécessaire, dans les conditions *abc*, une autre substance ou un autre événement, elle la cause, ou elle la détermine causalement.

Un événement peut être causé sans être nécessaire. Imaginons un individu dont le cerveau est relié à un appareil où surviennent des processus physiques indéterminés. Soit E un événement indéterminé, comme ceux qu'admet la physique quantique. Si l'événement E se produit dans l'appareil, notre cobaye choisit le vote républicain. Puisque l'événement E est contingent, le choix qui en résulte est également contingent, bien que l'individu désigné pour l'expérience n'en soit nullement responsable. Il aurait pu faire un autre choix, mais son choix a été causé par un événement qu'il ne contrôlait pas, et cela suffit pour qu'il ne soit pas responsable de son choix. La responsabilité morale d'une personne à l'égard d'un résultat suppose que celui-ci ne soit pas causé par quelque chose que cette personne ne contrôle pas.

Toute la difficulté est de savoir si l'absence de nécessité est une condition supplémentaire de la responsabilité morale. Je pense cependant qu'il n'est pas nécessaire de le

supposer et qu'en évitant de le supposer, on défend plus facilement la thèse incompatibiliste face à deux types de contre-exemples. Le premier a été imaginé par Harry Frankfurt²⁸ et repris maintes fois sous diverses formes. Dans toutes les variantes de cet exemple, l'agent est contrôlé ou limité de telle sorte que s'il choisit de faire A, quelque chose ou quelqu'un l'en empêche et le pousse à faire B, alors que s'il choisit de faire B, rien n'agit sur lui et il fait donc B de lui-même. Reprenons le cas imaginé par Frankfurt : Black est neurochirurgien. Jones, un de ses clients, doit être opéré à cause d'une tumeur au cerveau. Pendant l'opération, Black introduit dans le cerveau de Jones un mécanisme au moyen duquel il peut contrôler ses actions, et Jones ignore cela. Si Jones manifeste une inclination à voter pour Bush plutôt que pour Clinton, le mécanisme réagit en déclenchant le désir contraire, l'amenant ainsi à voter pour Clinton au lieu de Bush. Mais si, dans le cerveau de Jones, apparaît une inclination à voter pour Clinton, l'appareil ne réagit pas ; il le laisse agir normalement.

D'après Frankfurt, cet exemple imaginaire prouve qu'on peut être responsable d'un acte sans avoir la possibilité d'agir autrement. En effet, dans le second cas, où Jones manifeste une propension à voter pour Clinton, il agit selon sa propre volonté, et son vote n'est pas le résultat d'une cause agissant sur lui. Il est donc pleinement responsable de son acte, bien qu'il n'ait pas la possibilité d'agir autrement. Non seulement Jones agit par lui-même, mais nous pouvons parfaitement imaginer que sa volition n'est pas déterminée, elle non plus, par des causes extérieures. Dans cette hypothèse, Jones est libre et veut librement ; il exerce donc son libre arbitre, mais il ne peut pas agir autrement. Par conséquent, la liberté de l'acte ne serait pas nécessairement liée à l'existence de possibilités alternatives.

Dans tous les cas de ce genre, on peut se poser au moins deux questions :

Premièrement, est-ce que l'action de faire B *de soi-même* est la même que celle de faire B sous l'effet d'autre chose ? S'il s'agit de deux actions différentes, il ne s'agit pas d'un seul

²⁸ Harry Frankfurt, « Alternate Possibilities and Moral Responsibility », *Journal of Philosophy* 66 (1969), p. 329-839.

événement qui se produit dans tous les cas, donc il ne s'agit pas d'un événement nécessaire et inévitable.

Deuxièmement, en admettant que B soit une action et que l'agent n'ait pas de possibilité alternative à sa portée, on peut dire qu'il est responsable de B, mais cela ne milite pas en faveur du déterminisme. En effet, on peut considérer que s'il est responsable de B, c'est parce qu'il fait B de lui-même, et que s'il n'est pas responsable dans le cas où l'appareil intervient, c'est parce qu'il choisit sous l'influence d'une cause. Cette deuxième réponse n'est possible que pour quelqu'un qui fonde sa théorie incompatibiliste sur l'opposition entre la responsabilité et la nécessité *causale*. Pour celui qui oppose la responsabilité à la nécessité *en général*, il est crucial de montrer que dans tous les exemples à la Frankfurt, l'agent a la possibilité de ne pas faire l'action qu'il fait dans la situation actuelle.

Il existe une deuxième catégorie d'exemples que j'appellerai les exemples à la Dennett. Nous en avons déjà discutés un peu plus haut. Dans *Elbow Room*²⁹, Dennett dit à peu près ceci : je ne peux pas torturer quelqu'un pour une petite somme d'argent, ni même pour mille dollars. Si on veut m'y pousser, c'est librement et volontairement que je refuserai de le faire. On voit qu'ici, la nécessité excluant toute possibilité alternative n'est pas de nature causale, mais morale. On pourrait aussi l'appeler la *nécessité des raisons*, car elle liée à la rationalité de l'agent, à sa connaissance du Vrai et du Bien. En proposant cet exemple, Dennett pense aller plus loin que Frankfurt dans la défense du compatibilisme, mais son argument porte seulement sur la compatibilité de la *nécessité* avec le libre arbitre. Ni chez Frankfurt, ni chez Dennett, il n'est clairement question de nécessité *causale*. Je pense donc que ces exemples ne peuvent montrer que la compatibilité entre la nécessité et la liberté. Leurs arguments ne portent pas contre une hypothèse incompatibiliste fondée sur l'opposition entre la responsabilité et le fait d'être causé. Or cette hypothèse, au moyen du principe (NC), peut rendre compte de l'intuition libertarienne selon laquelle il y a une opposition entre la nécessité et la responsabilité morale. Une intuition dépend des exemples qui la confirment ou la mettent en évidence. Nous verrons dans les chapitres suivants de nombreux exemples utilisés par les

²⁹ Daniel Dennett, *Elbow Room*, 1984, p. 133.

libertariens pour montrer cette opposition entre la nécessité et la responsabilité morale. Puisqu'il est possible de rendre compte de tous ces exemples à partir d'un principe plus faible qui oppose la responsabilité avec la nécessité causale, je pense rien que n'interdit de se limiter à ce principe. Puisque mon but est seulement de montrer l'incompatibilité du déterminisme causal avec la responsabilité morale, je n'ai pas besoin d'une hypothèse plus forte. A l'inverse, j'essaierai de montrer dans les chapitres IV à VI qu'on ne peut pas se contenter d'un principe encore plus faible, compatibiliste ou semi-compatibiliste. Enfin, l'autonomie ultime est une condition qui inclut l'existence de possibilités alternatives, comme nous le verrons dans les chapitres II et III, alors que l'inverse n'est pas vrai : l'existence de possibilités alternatives, mêmes causales, n'inclut pas l'autonomie ultime, car l'agent pourrait être déterminé par un processus qu'il ne contrôle pas et qui est lui-même indéterminé.

1.11 Le restrictivisme et la liberté d'indifférence

Un autre argument pouvant militer en faveur de cette hypothèse est que si la liberté était liée essentiellement à la *possibilité* d'agir autrement, elle serait favorisée par l'indifférence. On sait que pour Descartes, « l'indifférence est le plus bas degré de la liberté ». A la suite de Descartes, beaucoup de philosophes ont opposé la liberté « éclairée » et la liberté d'indifférence. Celle-ci est l'état d'une personne qui a la possibilité d'agir de plusieurs façons mais qui ne voit aucune raison de préférer l'une à l'autre³⁰. Puisque les deux actions possibles (soit x et y) sont également probables, il semble que l'agent ait la même liberté d'accomplir x

³⁰ Dans les controverses théologiques du XVII^{ème} siècle sur la liberté et la grâce, on distinguait souvent deux sortes d'indifférence. Par exemple, le grand théologien jésuite Denis Petau (1583-1652) distingue « l'indifférence d'élection » (*electionis*) et « l'indifférence d'inclination » (*propensionis*). Selon Petau, sans une indifférence de la volonté par rapport aux causes, il n'y a pas de libre arbitre. Même une liberté éclairée suppose donc l'indifférence d'élection, sans laquelle il n'y a pas de responsabilité morale. Mais le libre arbitre et l'indifférence d'élection ne supposent pas l'indifférence d'inclination. Descartes lui-même, après avoir opposé la liberté d'indifférence et la liberté éclairée dans les *Méditations métaphysiques*, a dû changer de position (dans ses *Lettres à Mesland*), et reconnaître qu'il y a toujours une sorte d'indifférence dans la volonté, même lorsqu'elle est tellement éclairée qu'elle ne peut pas choisir le mal. J'admets pour ma part que la liberté suppose toujours le pouvoir d'agir ou de vouloir autrement, c'est-à-dire le fait de n'être pas déterminé causalement dans son action ou son vouloir (ou le fait de n'être pas empêché d'agir autrement). C'est en cela que consiste l'indifférence existant dans toute espèce de liberté métaphysique. Par contre, je pense que le libre arbitre ne suppose pas l'indifférence d'inclination. Sur la liberté d'indifférence dans les controverses théologiques, cf. Étienne Gilson, *La liberté selon Descartes et la théologie*, F. Alcan, Paris, 1913, et Olivier Boulnois, « Le refoulement de la liberté d'indifférence et les polémiques anti-scoistes de la métaphysique moderne », *Les Études philosophiques* n°2 (2002).

et y . C'est du moins une conséquence apparente de l'idée que la liberté se définit par la possibilité d'agir autrement. Si la probabilité d'une des actions est infime, on considère qu'il a à peine la possibilité de l'accomplir. Sa liberté de l'accomplir est donc quasiment nulle et on hésite à lui attribuer la capacité d'agir autrement (en supposant que x et y soient les seules possibilités ouvertes à l'agent). Il est vrai que chacune des deux actions reste possible tant que sa probabilité n'est pas nulle. Mais en disant cela, on ne répond pas de façon satisfaisante à l'objection, car elle s'appuie sur le fait qu'on puisse parler d'une liberté *plus ou moins grande*, donc de degrés de libertés.

Robert Kane estime que les probabilités initiales dans une situation de choix libre sont liées au caractère et aux motivations de l'agent, et peuvent se situer n'importe où entre 0 et 1, car elles ne déterminent pas le choix³¹. Mais le fait qu'elles ne le déterminent pas n'implique pas que l'agent ait une entière liberté à l'égard de x et de y . Ces probabilités peuvent indiquer que l'agent est fortement incliné vers une des possibilités, et cela peut être une circonstance atténuante, comme dans [superstition]. Si nous voyons le libre arbitre comme la forme de liberté que suppose la responsabilité morale, il est assez naturel d'admettre des degrés de liberté liés aux degrés de responsabilité. À partir du moment où un libertarien admet ce principe et suppose que la responsabilité morale est liée à la possibilité d'agir autrement, il doit pouvoir admettre que si la probabilité de x est très forte, l'agent n'a qu'une faible responsabilité en faisant x . La conclusion de ce raisonnement est que l'indifférence favorise la liberté puisqu'elle favorise l'équiprobabilité de x et de y .

Formulons l'argument de façon plus explicite. Le libertarien dont je parle admet le raisonnement suivant :

- (1) Si S fait x et que la probabilité de x avant le choix est égale à 1, alors S ne peut pas agir autrement.
- (2) S est responsable de l'action $x \rightarrow S$ a la possibilité d'agir autrement.
- (3) La probabilité que S fasse x est de 1.

³¹ Robert Kane, *The Significance of Free Will*, p. 177-178.

(4) Donc S ne peut pas agir autrement. (en vertu de (1) et (3))

(5) Donc S n'est pas responsable de l'action x . (en vertu de (2) et (4))

Or un libertarien qui admet ce raisonnement doit pouvoir admettre le raisonnement suivant :

(1) Si S fait x et que la probabilité de x avant le choix est très proche de 1, alors S peut à peine agir autrement.

(2) S est responsable de l'action $x \rightarrow$ S a la possibilité d'agir autrement.

(3) La probabilité que S fasse x est très proche de 1.

(4) Donc S peut à *peine* agir autrement.

(5) Donc S est à *peine* responsable de l'action x .

La conclusion du raisonnement est que pour que S soit pleinement responsable de x , la probabilité de x doit être à peu près égale à celle de y . Ainsi, la prémisse (2), selon laquelle la responsabilité morale suppose la possibilité d'agir autrement, aboutit à l'idée que l'indifférence favorise la liberté.

Pour admettre (2) sans accepter cette conséquence, il faut rejeter (1), c'est-à-dire refuser qu'il y ait des degrés entre « pouvoir faire x » et « ne pas pouvoir faire x ». Je ne vois pas comment justifier ce refus. La prémisse (2) entraîne donc vers l'idée qu'il y a un lien conceptuel entre le libre arbitre et l'indifférence : le libre arbitre suppose l'équiprobabilité, laquelle suppose l'indifférence. Il en découle une opposition entre le *libre arbitre* et la *liberté éclairée*, car s'il y a de bonnes raisons de faire x plutôt qu'autre chose, le fait de connaître ces raisons et d'y être sensible incline fortement vers l'action x et diminue d'autant la probabilité de faire autre chose.

Cette idée a des conséquences restrictivistes. On appelle *restrictivisme* la thèse selon laquelle nous faisons rarement usage du libre arbitre, même si nous agissons souvent de manière volontaire. Si le libre arbitre suppose l'indifférence, Dieu n'aurait pas le libre arbitre,

sauf lorsque deux possibilités sont également bonnes, et un homme très éclairé ne l'aurait que peu ou rarement. En augmentant sa connaissance du bien et du mal et en exerçant sa volonté à vouloir toujours le bien, on développerait sa liberté morale mais on limiterait l'usage de son libre arbitre³². Un individu parfaitement éclairé ne pourrait exercer le libre arbitre que dans les situations d'indifférence, où nulle possibilité n'est sensiblement meilleure que les autres ; il n'aurait donc le libre arbitre que pour des choix dépourvus de portée morale. Cette thèse n'est pas absurde, mais me paraît contraire à la manière dont nous employons les mots « liberté » et « libre arbitre ».³³

Laissons maintenant de côté la notion de libre arbitre, et concentrons-nous sur la responsabilité morale. Le point le plus important, après tout, est de savoir si la notion de responsabilité morale suppose la contingence au sens le plus strict. Si la responsabilité d'un agent S à l'égard d'une action A suppose que S ait la possibilité *au sens le plus strict* de ne pas faire A, il me semble qu'on doit tenir le raisonnement suivant :

- (1) Si S voit une bonne raison de faire A plutôt que de ne pas faire A³⁴, alors il est très probable que S choisisse de faire A.

³² Une thèse de ce genre se trouve chez Gibieuf, théologien qui a provoqué la controverse entre les jésuites et les jansénistes en déclarant, en 1630, que la thèse de Molina (jésuite) sur la liberté et la grâce avait été condamnée par Rome. Gibieuf pense que l'indifférence et l'évidence varient en proportion inverse : plus on sait ce qu'il faut faire, moins on a de liberté d'indifférence et plus on agit de façon éclairée. Gibieuf oppose l'indifférence et « l'amplitude », liberté consistant dans le détachement (et en ce sens l'indifférence) par rapport aux biens créés qui éloignent de Dieu. La thèse de Gibieuf a été approuvée par Jansenius. On a reproché à ces deux théologiens d'opposer la liberté morale et la liberté métaphysique. En effet, leur position signifie que l'acquisition de la liberté morale (l'amplitude de Gibieuf) supprime la liberté métaphysique (donc le libre arbitre). Cependant, une idée comme celle-ci n'implique pas qu'en développant sa liberté morale, on cesse d'être responsable de ses actions, car on peut être responsable de manière dérivée, ayant eu dans le passé la possibilité de s'orienter vers un genre de vie moins vertueux. Sur la liberté d'indifférence dans les débats théologiques, voir O. Boulnois, « Le refoulement de la liberté d'indifférence et les polémiques anti-scotistes de la métaphysique moderne », *Les Études philosophiques* n°2 (2002).

³³ Il me semble que nous appelons « libre arbitre » la liberté métaphysique que possède toute personne et qui lui permet d'être responsable de ses actions. Par « toute personne », j'entends non seulement les être humains normalement constitués, mais aussi une personne parfaitement intelligente ou omnisciente, comme Dieu. En cela, ma position a quelque chose de scotiste : je cherche une définition du libre arbitre qui s'applique de façon univoque à tous les êtres doués de raison et d'intelligence, quelles que soient les différences qui les séparent.

³⁴ Que S voie ou non des raisons de ne pas faire A. S peut avoir en effet à la fois des raisons de faire A et des raisons de ne pas faire A. Par exemple, il peut avoir peur de faire A (il met en jeu sa réputation, ou sa carrière), et penser en même temps qu'il doit prendre ce risque. Il peut aussi se trouver tiraillé entre un désir sensible et une

- (2) Sauf dans les cas de responsabilité dérivée, s'il est très probable que S choisisse de faire A, et que S fait effectivement A, alors S est à peine responsable de A.
- (3) Donc si S voit une bonne raison de faire A plutôt que de ne pas faire A, et fait effectivement A, alors (sauf s'il s'agit d'un cas de responsabilité dérivée), S est à peine responsable de A.
- (4) Donc pour que S soit pleinement responsable de A, il faut que la probabilité de l'événement « S fait A » et celle de l'événement « S ne fait pas A », soient proches.

Cela suppose une relative indifférence de l'agent à l'égard de ces deux possibilités. Or cette indifférence peut avoir trois causes. Soit elle vient de l'ignorance, soit elle vient de ce que l'agent est tiraillé entre sa connaissance du bien et un désir (sensible par exemple) qui n'obéit pas à cette connaissance, soit elle vient de ce que les deux comportements possibles sont réellement (ou objectivement) équivalents. Considérons ces trois genres de cas un à un :

1. **Indifférence liée à l'ignorance.** Un ami de Jean lui propose une drogue dangereuse, mais Jean ignore qu'elle est dangereuse. La probabilité que Jean accepte est à peu près égale à la probabilité qu'il refuse.
2. **Indifférence liée à un conflit de tendances.** Jean a déjà consommé de cette drogue. Il en garde un souvenir obsédant et a un puissant désir d'en reprendre. Mais, en même temps, il est parfaitement conscient du danger qui menace sa santé s'il en reprend. La probabilité que Jean accepte est à peu près égale à la probabilité qu'il refuse.
3. **Indifférence objective.** Il fait chaud et un ami de Jean lui propose une glace au melon. Jean aime ce parfum, mais n'a pas envie de manger une glace. Il y a environ une chance sur deux qu'il accepte, et les deux options sont moralement indifférentes

obligation morale. Dans la prémisse (1), je considère le cas d'une personne qui, grâce à un bon discernement (et quel que soit le soin qu'elle a pris à peser ces raisons) sait qu'elle doit faire A.

(il ferait peut-être plaisir à son ami en acceptant, mais il céderait peut-être aussi à la gourmandise). Finalement, il accepte.

Nous pouvons comparer ces exemples à un autre cas où il est très peu probable que Jean accepte la proposition qu'on lui fait : son ami lui propose une drogue dangereuse ; il sait qu'elle est dangereuse, et ne ressent aucun désir de consommer cette drogue, car il ignore les sensations qu'elle procure. La probabilité qu'il accepte est très faible mais pas nulle, car il pourrait goûter cette drogue par curiosité.

La question qui se pose maintenant est la suivante : l'indifférence peut-elle contribuer à la responsabilité morale de l'agent ? L'indifférence du premier genre est liée à l'ignorance. Elle ne peut donc contribuer à la responsabilité morale de l'agent. En effet, on n'est pas responsable de ce qu'on fait si on ignore la valeur de ce qu'on fait. A priori, l'indifférence du second type n'augmente pas non plus la responsabilité de l'agent par rapport à une situation où il choisit le bien sans tiraillement. Ici, nous pouvons distinguer deux cas de figure : celui où l'agent cède à son penchant mauvais, et celui où il résiste, grâce à sa connaissance du bien. Dans le premier cas, l'indifférence où il se trouve avant le choix n'augmente pas sa responsabilité puisque la présence d'un facteur qui exerce une pression sur la volonté de l'agent est normalement considérée comme une circonstance atténuante. Imaginons que Jean ait goûté cette drogue involontairement et en garde un souvenir obsédant qui le pousse à en reprendre. Le fait qu'il ait un tel désir constitue une circonstance atténuante : on l'excuse plus volontiers que s'il n'avait pas ce désir très fort et difficile à contrôler. Dans le deuxième cas de figure, où Jean résiste au contraire à ce désir, il semble avoir plus de mérite que s'il n'avait pas ce désir, mais sa responsabilité n'est pas plus grande, sinon la responsabilité à l'égard d'une action A supposerait l'inverse d'un pouvoir des alternatives : moins l'agent aurait le pouvoir de faire A, plus il serait responsable de A en faisant A.

Reste donc l'indifférence du troisième type. Or si la responsabilité morale suppose cette indifférence, il en découle que nous ne sommes responsables de nos actions ou de nos choix que lorsque l'enjeu moral est faible. La thèse selon laquelle la responsabilité morale s'oppose à la nécessité en général a ainsi des conséquences restrictivistes.

Elle semble aboutir à l'idée que si S choisit de faire A et fait A, S est responsable de A seulement si :

- (a) S est dans une relative indifférence à l'égard de A et non-A.
- (b) Et cette indifférence vient du fait qu'aucun de ces deux comportements n'est *réellement* meilleur que l'autre.

Cette conclusion ne s'applique pas aux cas de responsabilité dérivée, mais par définition, la responsabilité dérivée *dérive* d'un comportement antérieur de l'agent dont celui-ci était responsable d'une manière non-dérivée. Donc dans les cas de responsabilité dérivée, la responsabilité de l'agent à l'égard d'une action B suppose que B découle d'une action antérieure A dont S était responsable, et il ne pouvait être responsable de A que si les conditions (a) et (b) étaient remplies. Or les situations de choix où nous devons trancher entre deux possibilités quasiment équivalentes sont celles qui ont le moins de portée morale. La conséquence ultime de ce raisonnement est donc *qu'on n'est pleinement responsable moralement de son choix que dans les cas où il a une portée morale insignifiante*.

Je pense donc qu'il est prudent de ne pas supposer un lien conceptuel entre la responsabilité morale et la nécessité en général. Je défendrai la thèse incompatibiliste en soutenant l'idée que la nécessité causale, et plus largement, le fait d'être causé, exclut la responsabilité morale.

1.12 Résumé

Nous avons vu, au début de ce chapitre, l'importance du principe de crédulité : nous ne pouvons soutenir aucune idée vraisemblable sur les conditions de la responsabilité morale sans nous fier à nos « intuitions » spontanées, en essayant de les accorder entre elles. Refuser de tenir pour vrai ce qui paraît tel, comme le demande le principe de crédulité, c'est renoncer au fond à toute règle. Intuitivement, nous admettons qu'un agent est responsable de son action s'il l'accomplit volontairement, de manière autonome, et ayant eu la possibilité de ne pas l'accomplir. Mais aucune de ces trois conditions n'exclut de manière évidente la thèse déterminisme. Dans les sections 1.2 à 1.9, j'ai essayé de définir les termes de base les plus

importants tout en fixant dans ces définitions nos intuitions relatives aux pouvoirs et à la nécessité. En 1.2, j'ai distingué trois sens des mots « nécessaire » et « contingent ». En disant que la responsabilité suppose la possibilité d'agir autrement, les libertariens veulent au moins dire qu'un autre comportement est possible compte tenu à la fois des lois de la nature et du passé actuel. Ensuite (dans la section 1.4), j'ai montré qu'il fallait distinguer le pouvoir de la simple possibilité, puis j'ai distingué plusieurs espèces de capacités. Du point de vue libertarien, les seules possibilités alternatives pertinentes du point de vue de la responsabilité morale sont celles qui sont accessibles à l'agent, donc des possibilités qui sont en son pouvoir et ne relèvent pas du hasard, par exemple. Inversement, on peut se demander si la possibilité de faire *x* est une condition *nécessaire* du pouvoir de faire *x*. A partir de la section 1.5, j'ai présenté l'hypothèse selon laquelle ce n'est pas une condition nécessaire de la responsabilité morale, ni du pouvoir de faire *x*. En 1.6, j'ai distingué trois espèces de capacités, pour montrer qu'aucune ne suppose *a priori* des possibilités alternatives au sens large, et que l'acquisition des facultés de la troisième espèce (les dispositions) diminue même la possibilité (ou la probabilité) des actions contraires à ces dispositions. Dans les sections suivantes (1.7 à 1.9), j'ai donc soutenu que ce n'est pas la nécessité *en général* qui s'oppose à la liberté et à la responsabilité morale, mais seulement la nécessité *causale*. Le degré de responsabilité ne varie pas systématiquement en fonction de la probabilité des possibilités alternatives. Il faut donc distinguer les probabilités qui sont dues à une influence causale et les autres probabilités (dues par exemple à une disposition acquise par l'agent). Seules les premières, à mon avis, diminuent la liberté et la responsabilité de l'agent. En 1.10, j'ai montré que cette hypothèse rend la thèse incompatibiliste moins vulnérable aux objections compatibilistes. Enfin, en 1.11, nous avons vu que si la liberté supposait la possibilité (et non seulement la possibilité *causale*) d'agir autrement, cela signifie qu'il y a un lien conceptuel entre le libre arbitre et l'indifférence et - ce qui est plus gênant - un lien conceptuel entre la responsabilité morale et l'indifférence.

Dans les chapitres II à VI, je critiquerai donc les thèses déterministes et compatibilistes sans supposer que la nécessité *en général* s'oppose à la responsabilité morale. Je m'appuierai seulement sur l'idée que la nécessité *causale*, donc le déterminisme causal, exclut la liberté et la responsabilité morale. Dans le chapitre II, j'examinerai les arguments pour et contre le principe des possibilités alternatives. Nous verrons qu'il est difficile de le défendre sans le

reformuler. Ensuite (dans le chapitre III), je défendrai un deuxième principe à mon avis plus efficace pour critiquer le déterminisme : le principe du transfert.

II La responsabilité et le pouvoir des alternatives

Une part considérable des discussions récentes ou actuelles sur le déterminisme et la responsabilité morale, en philosophie analytique, porte plus particulièrement sur ce qu'on appelle le *principe des possibilités alternatives*. Sous sa forme la plus simple, il s'énonce de la manière suivante :

(PAP) Pour tout agent S et toute action A, si S accomplit l'action A, il est responsable de A seulement si, avant de faire cette action, il avait la possibilité de ne pas la faire.

PAP exprime une idée que la plupart des gens, au moins dans la culture occidentale, admettent spontanément et que beaucoup utilisent comme critère pour savoir si un malfaiteur est responsable de son méfait. Par exemple, si un individu a commis un crime sous l'effet d'une pulsion à laquelle il pouvait difficilement résister, nous disons que cette influence est une circonstance atténuante, et que cet individu est faiblement responsable de son action. Or Harry Frankfurt et d'autres philosophes ont présenté des contre-exemples pour réfuter ce principe. Ces contre-exemples sont devenus depuis un des principaux fondements de la thèse compatibiliste. Dans ce chapitre, je procéderai à une étude systématique de ces exemples pour les discuter et pour évaluer le principe des possibilités alternatives.

Dans un exemple classique à la Frankfurt, l'agent fait une action A de lui-même, mais s'il voulait faire autre chose, un facteur interviendrait pour lui faire choisir A. J'ai déjà présenté un tel exemple plus haut :

- Gunnar est contrôlé par un neurochirurgien qui a installé un appareil dans son cerveau pour influencer son comportement. Si à t1, Gunnar ne manifeste aucun désir de tirer sur Ridley à t2, l'appareil produit en Gunnar le désir de tirer sur Ridley à t2.
- A t1, Gunnar sort son pistolet de sa poche pour tirer sur Ridley. A t2, il tire sur Ridley, qui meurt aussitôt après. Puisque Gunnar a fait cela de lui-même, l'appareil

n'est pas intervenu. Gunnar est donc responsable de ce qu'il fait, même s'il ne pouvait pas éviter de tuer Ridley.

En 2.1, je reprendrai cet exemple et je présenterai la réponse la plus naturelle : celle du « frémissement de liberté » (*flicker of freedom*), qui s'appuie sur le fait que même si l'agent ne peut agir autrement, il peut encore choisir autrement. En 2.2, je discuterai un exemple à la Frankfurt où l'agent n'a même pas la possibilité de choisir autrement. Je soutiendrai que dans cet exemple inventé par Eleonore Stump, l'agent n'est pas responsable de son choix. En 2.3, je répondrai à cet exemple par l'argument du « dilemme » utilisé par Widerker pour neutraliser tous les exemples à la Frankfurt. En 2.4, je discuterai un contre-exemple imaginé par Widerker lui-même pour répondre à sa propre objection. Je montrerai que la cohérence de cet exemple est douteuse. Dans leurs contre-exemples, Stump et Widerker supposent que l'agent est responsable si son action n'est pas déterminée causalement. Ainsi, même si l'action découle d'un processus physique indéterministe non-contrôlé par l'agent, celui-ci serait responsable de son action. Or ce présupposé est douteux. Il est permis de croire que l'indéterminisme physique imaginé à la source du choix ne rend pas l'agent plus libre et plus responsable de son choix que s'il était causalement déterminé. A ce point de la discussion, le compatibilisme se heurte au dilemme suivant. Dans tous les exemples étudiés jusque là, même si l'agent ne peut faire un autre choix que son choix actuel, il y a des possibilités alternatives (au sens large) à la source du choix : soit elles sont accessibles à l'agent, soit elles ne lui sont pas accessibles. Dans le premier cas, un libertarien peut soutenir que l'agent a encore la possibilité de vouloir autrement (c'est pourquoi il est responsable) ; et dans le second cas, le libertarien peut refuser l'exemple en disant que l'agent n'est pas responsable.

Pour échapper à ce dilemme, la première solution est d'imaginer un exemple de blocage, où l'agent n'a plus les moindres possibilités alternatives. J'examinerai cette stratégie de Mele, Robb, et Hunt dans la section 2.6. Après l'avoir critiquée, je discuterai enfin (en 2.7) la dernière stratégie possible. Si les exemples de blocage ne fonctionnent pas, la dernière possibilité est de revenir à des exemples où il y a des possibilités alternatives, mais en montrant que les possibilités alternatives en question ne sont pas « moralement pertinentes ». C'est la stratégie de McKenna et Pereboom, qui énoncent des critères de pertinence morale

pour montrer que dans certains cas, l'agent est responsable de son action alors que le « frémissement de liberté » situé à la source de l'acte est négligeable.

2.1 Les exemples à la Frankfurt

Il existe des situations où un individu se rend lui-même incapable de ne pas accomplir une certaine action. Par exemple, un homme qui se rend lui-même dépendant de la drogue à partir de t_1 peut se trouver ensuite dans une situation où il perd le contrôle de lui-même et frappe quelqu'un à t_2 . En supposant qu'il n'ait pas la capacité de se retenir, cet exemple contredit alors PAP. Mais il s'agit visiblement d'un cas de responsabilité dérivée. En effet, si l'agent n'était pas lui-même la cause de son incapacité, on ne le jugerait pas responsable de son acte. Un incompatibiliste dira alors que s'il est bien responsable moralement, c'est parce qu'il a eu, dans une situation antérieure, la possibilité de faire un choix ou une action lui permettant de ne pas frapper quelqu'un à t_2 . En effet, il aurait pu ne pas se droguer à partir de t_1 . Ainsi l'agent est responsable à t_2 parce que sa situation à un moment antérieur était conforme à PAP. Je pense donc que les cas de responsabilité dérivée sont à considérer comme des exceptions qui ne réfutent pas vraiment PAP. En revanche, les exemples dits « à la Frankfurt » constituent une objection sérieuse contre ce principe.

Frankfurt¹ lui-même a proposé l'exemple suivant : Jones va voter. Or Black, un neurochirurgien, contrôle ses choix au moyen d'un appareil installé dans le cerveau de son client. Si l'appareil détecte une inclination à voter pour Bush, il provoque aussitôt le désir de voter pour Clinton. Mais si les états du cerveau de Jones laissent deviner une tendance à voter pour Clinton, l'appareil ne réagit pas : il laisse Jones voter par lui-même. Cet exemple est censé prouver que la notion de responsabilité est compatible avec l'absence de possibilité alternative. Dans tous les cas où Jones agit *par lui-même*, sans être influencé par une cause extérieure, nous le considérons comme responsable de ses actes. En l'occurrence, si Jones

¹ Harry Frankfurt, « Alternate Possibilities and Moral Responsibility », *Journal of Philosophy* 66 (1969), p. 823-839.

vote par lui-même pour Clinton, sans intervention de la machine, il est responsable de son action, quoiqu'il n'ait pas la possibilité d'agir autrement.

Jones ne peut pas éviter de voter pour Clinton, mais nous pouvons remarquer qu'il a tout de même des possibilités alternatives dans l'orientation de sa volonté. En effet, même si Jones ne peut pas voter pour Bush, il peut quand même avoir une *inclination* vers ce choix. Il ne peut pas agir autrement, mais il peut encore vouloir autrement. Il y a donc des possibilités alternatives à la source de l'acte. Cette thèse est souvent appelée la thèse du « frémissement de liberté » (« *flicker of freedom* »). Elle propose de remplacer PAP par un principe de ce genre:

(PFL) (principe du frémissement de liberté) : Si S accomplit l'action A, il est responsable de A seulement si, avant de la faire, il avait la possibilité de ne pas la faire ou de ne pas vouloir la faire.

Fischer répond à cette thèse en disant que la possibilité de vouloir autrement n'est pas significative du point de vue moral. Pour lui, la condition principale de la responsabilité est à chercher ailleurs : l'agent doit exercer un contrôle rationnel de ses propres actions. En agissant selon sa raison, l'agent est autonome, mais il n'est pas pour autant une cause première. Dans *The Metaphysics of Free Will*, Fischer distingue deux conceptions de la manière dont l'agent contrôle ses propres actes : le contrôle de régulation (« *regulative control* »), et le contrôle de direction (« *guidance control* ») :

Imaginons que je conduise ma voiture. Elle fonctionne bien et je veux faire un virage à droite. En conséquence de mon intention de tourner à droite, je clignote, tourne le volant, et dirige prudemment ma voiture vers la droite. En outre, je suppose ici que j'étais capable de former l'intention de ne pas tourner à droite, mais plutôt à gauche (...). Dans ce cas ordinaire, je dirige ma voiture vers la droite, mais j'aurais pu la diriger vers la gauche. Je contrôle la voiture, et j'exerce aussi un certain contrôle *sur* les mouvements de la voiture. Dans la mesure où je dirige effectivement la voiture d'une certaine manière, je peux dire que j'ai un « contrôle de direction ». En outre, dans la mesure où j'ai le pouvoir de diriger la voiture d'une autre manière, je peux dire que j'ai un « contrôle de régulation ».²

² John Martin Fischer, *The Metaphysics of Free Will*, Blackwell, Oxford, 1994, p. 132.

Ensuite, Fischer imagine une autre situation :

Le volant de la voiture fonctionne correctement si je dirige la voiture vers la droite. Mais à mon insu, le volant de la voiture est cassé, de telle sorte que si j'essayais de tourner dans une autre direction, la voiture virerait à droite exactement comme dans la situation présente.

Dans ce cas, il n'y a pas de possibilité alternative et le conducteur n'exerce aucun contrôle de régulation, mais il a tout de même un contrôle de direction (« *guidance control* ») sur son véhicule. D'après Fischer, c'est ce contrôle de direction, et lui seul, qui importe du point de vue de la responsabilité morale. Elle ne suppose pas la capacité d'agir autrement.

2.2 Le principe de prévention possible (PPP)

L'interprétation des exemples à la Frankfurt est souvent loin d'être évidente. Une première réponse est que dans ces exemples, si l'agent ne peut pas *agir* autrement, il peut tout de même *choisir* autrement. Or il est possible d'aller beaucoup plus loin. En effet, Van Inwagen a même montré qu'il est possible de refuser la conclusion de Frankfurt selon laquelle dans cette catégorie d'exemples, l'agent n'a plus la possibilité d'agir autrement³.

La stratégie de Van Inwagen consiste à reformuler le principe selon lequel la responsabilité morale suppose la capacité d'agir autrement. Il montre que PAP peut être remplacé par une autre version du principe qui ne tombe pas sous le coup des contre-exemples à la Frankfurt. Ce nouveau principe (« principe de prévention possible »), Van Inwagen le présente sous deux formes successives :

Je vais poser un certain principe concernant l'absence de responsabilité qui me semble être plausible, si les événements ou états de choses dont nous tenons les gens pour responsables sont des particuliers ; et je vais poser un principe similaire qui me semble être plausible si les événements ou états de choses dont nous tenons les gens pour responsables sont des universaux. Pour chacun de ces principes, je vais soutenir qu'il ne peut être réfuté par des contre-exemples dans le style de Frankfurt.

³ Peter Van Inwagen, *An Essay on Free Will*, Oxford University Press, Oxford, 1983, p. 83 et s.

Le premier de ces principes (« *principles of possible prevention* ») est :

(PPP 1) Une personne n'est moralement responsable d'un certain événement particulier que si elle pouvait l'empêcher.

Pour appliquer cette idée, il faut savoir ce qu'est un événement particulier. Il faut donc trouver un critère efficace pour savoir dans quels cas deux événements sont différents, et dans quels cas ils sont identiques. Van Inwagen reprend un critère qui vient de Davidson :

(déf.) Si x et y sont des événements, alors $x = y$ si et seulement si x et y ont les mêmes causes et les mêmes effet.

Van Inwagen propose d'admettre au moins la première condition : $x = y$ si et seulement si x et y ont les mêmes causes. Il observe que nous avons l'habitude d'appliquer ce critère pour définir l'identité d'une personne. En effet, on admet communément que deux personnes x et y sont identiques si elles ont la même origine causale. Il faut donc qu'elles viennent de la même cellule-œuf et des mêmes gamètes. Mais il faut aussi qu'elles soient numériquement identiques, car les jumeaux viennent de la même cellule-œuf et n'en sont pas moins des personnes différentes. Il faut donc que leur passé soit *exactement* le même. Tout cela revient à considérer deux personnes comme identiques si elles ont la même histoire causale. En appliquant cela aux événements, on obtient le critère suivant :

(déf.) « Si x est le produit de certaines causes, alors nécessairement, un événement y est le produit de ces causes si y est x ». ⁴

En d'autres mots, « aucun événement n'aurait pu avoir d'autres causes que ses causes actuelles ». Or, dans les exemples à la Frankfurt, les deux événements envisagés n'ont jamais la même histoire causale. Par exemple, quand Jones vote pour Clinton sous l'influence de la machine, l'événement « Jones vote pour Clinton » n'a pas la même histoire causale que dans le cas où il vote aussi pour Clinton, mais de lui-même, sans y être poussé par la machine. Van Inwagen imagine un exemple analogue : Gunnar tire et tue Ridley intentionnellement. Il cause

⁴ *Ibid.*, p. 170.

donc un certain événement : la mort de Ridley. Or s'il ne voulait pas tuer Ridley, il y aurait un certain facteur qui agirait sur lui de telle sorte qu'il finirait par tuer Ridley. Dans la situation actuelle, ce facteur n'intervient pas et Gunnar tue Ridley par lui-même. Il semble d'abord que Gunnar soit responsable de la mort de Ridley, malgré l'absence de possibilité alternative. Selon Van Inwagen, Gunnar est bien responsable, mais parce qu'il a véritablement une possibilité alternative. En effet, selon le critère d'identité formulé plus haut, l'événement que dénote l'expression « la mort de Ridley » n'est pas le même dans les deux cas, le facteur extérieur n'intervenant évidemment que dans le premier cas. Aussi, puisque l'histoire causale de l'événement « la mort de Ridley » n'est pas la même, il ne s'agit pas du même événement, mais de deux événements distincts. Finalement, on peut dire que Gunnar a des possibilités alternatives, non seulement au sens où il peut avoir plusieurs inclinations différentes, mais au sens où il peut causer deux événements différents.

Cette conclusion paraît sans doute un peu factice, mais elle s'impose à partir du moment où l'on admet que les événements dont on est responsable en les causant sont toujours des particuliers. Pour éviter cette conclusion, il n'y a pas d'autre issue que de considérer les événements dont nous parlons comme des universaux. Mais alors, les tenants de l'incompatibilisme peuvent abandonner PPP1 pour argumenter au moyen d'une deuxième version du principe :

(PPP2) « Une personne n'est moralement responsable d'un certain état de choses que si (cet état de choses se produit et si) elle pouvait l'empêcher de se produire ».⁵

On suppose que l'état de choses dont il s'agit est un universel. Or un état de choses universel peut se produire de plusieurs façons. L'état de choses « César est assassiné » s'est réalisé en 44 av. JC, à Rome, quand César fut poignardé, mais il aurait pu⁶ se réaliser autrement, par exemple si Cléopâtre avait emprisonné César à Alexandrie en 48 av. JC. On voit qu'un même état de choses universel peut avoir différentes origines causales. Il semble donc que si nous interprétons le principe des possibilités alternatives comme ici, en PPP2, il

⁵ *Ibid.*, p. 171.

⁶ Par exemple, c'était logiquement possible, et compatible avec les lois de la nature.

est de nouveau possible de réfuter ce principe au moyen des contre-exemples à la Frankfurt. Par exemple, si « la mort de Ridley » dénote un événement universel, il est juste de dire que Gunnar ne pouvait pas l'empêcher de se produire. Il pouvait tuer Ridley au terme d'une autre histoire causale, mais il ne pouvait pas éviter de tuer Ridley, et il n'avait donc pas la possibilité d'agir autrement, c'est-à-dire d'amener un autre événement. Mais en réalité, cette objection n'est pas valable. Car si « la mort de Ridley » est un universel, il est évident, non seulement que Gunnar ne peut pas l'empêcher, mais aussi qu'il n'en est pas responsable, même s'il n'est influencé par aucune cause. En effet, il est inévitable que Ridley meure un jour, comme tout être humain, et Gunnar ne peut être tenu pour responsable de sa mortalité. Considérons l'état de choses « Ridley est mortel » :

Quand Gunnar tire sur Ridley, il accomplit un acte suffisant pour la réalisation de cet état de choses. Mais il serait absurde de tenir Gunnar pour responsable de C (Ridley est mortel). On peut tenir Dieu, ou Adam et Eve ensemble (ou peut-être personne) pour responsables de la mortalité de Ridley ; mais certainement pas son meurtrier. (A moins, bien sûr, que Ridley eût toujours continué à vivre s'il n'eût été assassiné, mais admettons que ce ne fût pas le cas).⁷

D'après Van Inwagen, si l'état de choses en question (« Ridley meurt ») peut être considéré comme un universel, alors il est identique à l'état de choses « Ridley est mortel ». Or il est évident que Gunnar n'est pas responsable de la mortalité de Ridley. On ne peut donc pas s'appuyer sur l'exemple de Gunnar et Ridley pour déclarer que Gunnar est responsable de cet état de choses sans avoir la possibilité de l'éviter, car Gunnar n'est justement pas responsable de l'état de choses en question. En somme, l'exemple ne prouve pas que la responsabilité soit compatible avec l'absence de possibilité alternative. Ainsi, PPP2 est inattaquable si on considère les événements comme des universaux, et PPP1 est inattaquable si on les considère comme des particuliers. Dans un cas comme dans l'autre, il y a donc une formulation du principe des possibilités alternatives qui résiste aux contre-exemples à la Frankfurt. Apparemment, il faut conclure que la responsabilité morale suppose l'existence de possibilités alternatives.

⁷ *Ibid.*, p. 172.

Cependant, une étape de ce raisonnement peut sembler contestable : l'état de chose A (Ridley meurt), même défini comme un universel, est-il vraiment identique à l'état de choses B (Ridley est mortel) ? Être mortel, c'est se trouver dans une condition telle qu'on ne peut pas ne pas mourir un jour. « Ridley est mortel » signifie donc « il est nécessaire que Ridley meure », et non pas « Ridley meurt ». Du coup, il est normal que nous ne tenions pas Gunnar pour responsable de ce fait. On peut considérer Gunnar comme responsable de la mort de Ridley, mais il n'est pas responsable de la *nécessité que Ridley meure*.

À mon avis, cette objection ne remet pas en cause l'argument de Van Inwagen, mais plutôt la manière dont il présente l'exemple. Laissons de côté la notion de mortalité, et considérons l'événement « Ridley meurt ». Si cet événement est un universel, on ne peut pas en être responsable en le causant. Être responsable d'un fait en le causant, c'est toujours être responsable d'un fait particulier, d'un *événement-token*, et non d'un *événement-type*. Assurément, ce qu'on reproche à Gunnar, c'est d'avoir tué Ridley, et non de l'avoir tué de telle ou telle manière, mais l'événement dont il est responsable est bien un événement singulier ; et il n'y a pas d'universel dans le monde des causes et des effets, mais seulement des particuliers, qu'il s'agisse de substances, de propriétés, ou d'événements. Nous utilisons des universels pour décrire le monde de la causalité, mais il ne comprend que des particuliers. Si ce principe ontologique est correct, Gunnar n'est pas responsable de la mort de Ridley envisagée comme un universel.

Enfin, nous pouvons ajouter que si l'événement « Ridley meurt » pouvait être compris comme un universel, il faudrait sans doute en attribuer la responsabilité à l'intervenant contrefactuel, car ce n'est pas Gunnar, mais l'intervenant contrefactuel qui garantit le résultat « Ridley meurt » compris comme un universel. L'intervenant contrefactuel ne garantit pas (ne détermine pas) l'événement *particulier* qui se produit dans la situation actuelle. Il n'en est donc pas responsable. Mais en revanche, c'est bien lui qui rend causalement impossible tout résultat ne tombant pas sous la description *générale* « Ridley meurt » ; c'est donc lui qui détermine l'événement-type « Ridley meurt ». Ainsi, s'il existe un *événement-type* « Ridley meurt » et s'il est possible d'être responsable d'un *événement-type*, Gunnar n'en est pas responsable. Dans cette hypothèse, les contre-exemples à la Frankfurt manqueraient leur objectif, puisqu'ils ne sont pas des cas de responsabilité morale. Pour ma part, j'estime que

dans un cas frankfurtien classique, comme celui-ci, l'agent paraît au contraire responsable de ce qu'il fait. L'hypothèse qui précède n'est donc pas correcte : la responsabilité à l'égard d'un événement ne peut être comprise comme la responsabilité à l'égard d'un universel (un événement-type), sinon l'agent ne serait pas responsable de ce qu'il fait dans les exemples à la Frankfurt.

La réponse de Van Inwagen, ainsi complétée, me paraît valable. L'événement « Ridley meurt » est en fait un particulier, et si l'intervenant contrefactuel déclençait la mort de Ridley en agissant sur les pensées de Gunnar, il ne produirait pas le même événement. Il y a donc deux événements possibles : « Gunnar tue Ridley par lui-même », et « Gunnar tue Ridley sous l'influence de l'intervenant contrefactuel ». Les exemples à la Frankfurt sont des situations où un individu S ne peut pas éviter de faire une action tombant sous une certaine description générale mais où l'agent possède tout de même une capacité des alternatives, et même une capacité d'agir autrement. Le « frémissement de liberté » qu'il possède à la source de son action suffit à rendre possibles deux actions différentes.

Dans les prochaines sections, et jusqu'à la fin de ce chapitre, je discuterai plusieurs objections contre le principe des possibilités alternatives qui cherchent à exclure les réponses fondées sur l'existence d'un « frémissement de liberté » à la source de l'acte. Pour écarter ces réponses, les compatibilistes ont cherché des exemples dans lesquels il n'y a pas les moindres possibilités alternatives, ou dans lesquels les possibilités alternatives qui subsistent encore ne sont pas « moralement pertinentes ».

2.3 Le frémissement de liberté : un résidu des possibilités alternatives

On appelle stratégie du « frémissement de liberté » (« *flicker of freedom* ») la stratégie consistant à défendre le principe des possibilités alternatives en arguant que dans tous les exemples à la Frankfurt, il y a au moins un commencement de possibilités alternatives. La réponse de Van Inwagen que nous venons d'examiner peut être vue comme appartenant à ce genre d'argument. D'après Fischer, la stratégie du frémissement de liberté ne permet pas de sauver le principe des possibilités alternatives, car les possibilités qui restent ouvertes à

l'agent dans les situations à la Frankfurt ne sont pas suffisantes ; elles sont d'une portée trop limitée pour avoir une pertinence du point de vue de la responsabilité morale.

Dans les prochaines sections, j'examinerai cette question en passant en revue les différents types d'objections contre la stratégie du *flicker of freedom*. Certains auteurs comme Eleonore Stump ont imaginé des contre-exemples où l'agent n'a pas la possibilité de choisir autrement, l'intervenant contrefactuel ayant les moyens de deviner ses choix à l'avance. Ce genre de cas pose plusieurs problèmes, en particulier parce qu'on peut encore concevoir des possibilités alternatives antérieures au choix. Il se heurte également à l'argument du dilemme, utilisé par Kane et Widerker. Pour éviter cet argument, Mele et Robe ont imaginé des contre-exemples où il n'y a pas d'intervenant contrefactuel, ni d'indice permettant à celui-ci d'anticiper les choix de l'agent. Mais nous verrons que ces exemples posent encore problème, parce qu'ils supposent des possibilités alternatives avant le choix. Une solution radicale est alors d'imaginer un « blocage parfait », sans intervenant contrefactuel, ni indice, ni possibilité alternative d'aucune sorte (c'est la stratégie adoptée par David Hunt). Or les exemples de blocage parfait ont un défaut majeur : ils semblent impliquer un « déterminisme pur et simple », comme le dit Kane. Une dernière solution compatibiliste est de revenir à un exemple de blocage « limité », comme chez Mele et Robb, mais en montrant plus clairement que les seules possibilités alternatives supposées par l'exemple n'ont rien à voir avec la responsabilité morale de l'agent. C'est la stratégie adoptée par McKenna.

Comme Fischer, Eleonore Stump⁸ exprime d'abord des doutes sur la nature de la possibilité alternative sur laquelle repose l'argument du frémissement de liberté : *ou bien* il ne s'agit pas vraiment d'une possibilité alternative, *ou bien* nous devons reconnaître que c'en est une. Mais dans la deuxième hypothèse, on peut réfuter l'argument du frémissement de liberté en imaginant des contre-exemples à la Frankfurt où l'agent est moralement responsable alors qu'il n'a même pas la possibilité de *choisir* autrement. Ainsi, dans chacune des branches du dilemme, l'argument est réfuté. Telle est la forme de l'objection. Voyons maintenant les deux branches du dilemme.

⁸ Eleonore Stump, "Alternative Possibilities and Moral Responsibility : the Flicker of Freedom" (1999), p. 301.

D'après Stump, la stratégie du frémissement de liberté « requiert la supposition que faire une action-de-soi-même soit déjà une action à part »⁹. C'est en effet le cas si on admet que tuer Ridley par soi-même n'est pas la même action que tuer Ridley sous l'effet d'une cause incontrôlable. Nous retrouvons ici une idée énoncée par Van Inwagen : des événements décrits de la même façon sont néanmoins différents s'ils n'ont pas la même histoire causale. Or cette idée, Stump la trouve suspecte.

En admettant maintenant qu'elle soit recevable, elle n'est pas suffisante car nous pouvons imaginer des contre-exemples où il n'y a pas les moindres possibilités alternatives :

(G) Supposons qu'un neurochirurgien (Grey) veuille que Jones, son patient, vote pour les républicains aux prochaines élections ; Grey a un neuroscope lui permettant à la fois d'observer et de provoquer l'activation des neurones qui sont liés aux actes de volonté de Jones. A travers son neuroscope, Grey observe qu'à chaque fois que Jones veut voter pour les candidats républicains, cet acte de volonté est corrélé à l'achèvement d'une série d'activations de neurones dans le cerveau de Jones, et qu'elle inclut toujours, près de son commencement, l'activation des neurones a, b et c (appelons « R » cette séquence neuronale). D'autre part, la volonté de Jones de voter pour les candidats démocrates est corrélée à l'achèvement d'une autre séquence neuronale incluant, près de son commencement, l'activation des neurones x, y, z, dont aucun n'est identique à un neurone de la séquence R (appelons « D » cette séquence neuronale).¹⁰

On devine la suite de l'histoire : si Jones veut voter pour le candidat démocrate, le neurochirurgien l'en empêche, mais ce qui est encore mieux, c'est que le dispositif lui permet d'éviter que Jones ait la *volonté* de le faire, même un bref instant. En effet, cette volonté n'existe pas à proprement parler tant que la séquence neuronale n'est pas terminée. Aussi Grey peut-il régler l'appareil de manière à faire avorter toute séquence neuronale commençant par l'activation de x. L'appareil empêchera que se produise la séquence x-y-z et déclenchera la séquence a-b-c pour que Jones vote républicain. Il s'ensuivra une série d'événements que nous pouvons représenter ainsi : x | a-b-c.

⁹ *Ibid.*, p. 301.

¹⁰ *Ibid.*, p. 303.

Imaginons maintenant que Jones vote de lui-même pour le candidat républicain. Le neurochirurgien n'a pas besoin d'intervenir puisque le résultat qu'il attend est sur le point de se réaliser. Il n'intervient donc pas, et Jones est entièrement responsable de son vote. Pourtant, il n'a ni la capacité d'agir autrement, ni même celle de choisir autrement.

L'objection de Stump est très radicale, parce qu'elle remonte non seulement à la source de l'acte, mais à la source du choix. Seulement, une telle expérience de pensée dépend de plusieurs hypothèses :

1° - Stump suppose qu'un choix ou une volition est réalisée dans (ou « corrélée à ») une séquence d'événements physiques. Si le choix correspondait uniquement au terme de la série, c'est-à-dire c ou z, il serait déterminé par les événements antérieurs de la série. Jones ne serait donc pas libre, ni responsable de son vote, et l'exemple de Stump pourrait être refusé. Si à l'inverse, le choix correspondait au premier événement de la série, il serait vraiment indéterminé, mais imprévisible, et Grey ne pourrait donc pas intervenir à temps. Là non plus, l'exemple de Stump ne serait pas valable. Pour qu'il le soit, il est impératif que le choix soit corrélé à une série d'événements physiques. Comme il n'y a pas de demi-choix (ou de demi-volitions), le choix en lui-même est *un* (indivisible) et la relation entre ce choix et la séquence neuronale est une relation un-plusieurs (« *one-many relation* »)¹¹. Cette idée est généralement associée à l'émergence. William Hasker¹² et ceux qui défendent une théorie de l'émergence soutiennent qu'un état psychique *émerge* à partir d'une structure d'éléments matériels tout en ayant une unité forte, que l'on ne peut comprendre en analysant cette structure telle qu'elle se présente en elle-même. L'expérience de pensée imaginée par Stump présuppose-t-elle donc une conception émergentiste de l'esprit humain ? D'après elle, non ; elle est compatible avec différentes sortes d'émergentismes ou de matérialismes, et n'exclut que le « dualisme cartésien ». Un matérialiste « dur » pourrait, contrairement à Stump, admettre la possibilité de

¹¹ *Ibid.*, p. 305-306.

¹² Cf. William Hasker, *The Emergent Self*, Cornell University Press, Ithaca, 1999. Timothy O'Connor soutient aussi une forme d'émergentisme. Cf. Timothy O'Connor, *Persons and Causes*, Oxford University Press, Oxford, 2000.

demi-choix ou de choix incomplets, ce qui ferait échouer l'argument. Mais s'il refuse lui aussi cette possibilité, alors il admettra comme Stump une relation un-plusieurs entre le choix et sa réalisation physique. Le dualisme n'étant plus très répandu, Stump peut être sûre que la plupart de ses lecteurs lui accorderont cette hypothèse. Dans un article ultérieur¹³, Stump essaie de montrer pourquoi d'un point de vue libertarien, on ne peut pas refuser son idée d'une corrélation un-plusieurs autrement qu'en s'appuyant sur une hypothèse dualiste.

Voici son raisonnement. Soit le libertarien admet l'hypothèse de Stump, soit il la refuse en affirmant que la relation entre le choix et sa réalisation physique dans le cerveau est une relation un-un. Dans cette seconde hypothèse, l'événement physique simple auquel correspond le choix ne peut être qu'un événement physique indéterminé (a ou x dans notre exemple). En effet, s'il s'agissait d'un événement postérieur (comme c ou z), il serait déterminé par un événement antérieur, lui-même indéterminé. Il faut donc que le choix corresponde au tout premier événement de la série, le seul qui soit indéterminé. Or, déclare Stump, « dans aucune théorie raisonnable sur les relations entre l'esprit et le cerveau chez l'être humain, un acte mental de quelque nature n'est corrélé à un simple événement quantique »¹⁴. Le libertarien ne peut donc pas soutenir que le choix soit corrélé seulement à a (ou à x). Il ne lui reste qu'une seule possibilité : dire que le choix n'est réalisé ni dans a, ni dans c, ni dans a-b-c, mais qu'il est de nature purement mentale ; ainsi il n'aura pas de réalisation physique, mais sera tout au plus en interaction avec des états physiques du cerveau. Nous voyons ainsi quel est le dualisme « cartésien » auquel Stump faisait allusion.

Peu de philosophes, aujourd'hui, sont prêts à défendre le dualisme cartésien. Mais celui dont parle Stump n'est pas autre chose que le dualisme *en général*¹⁵. Descartes suppose que

¹³ E. Stump, "Moral Responsibility without Alternative Possibilities", in *Moral Responsibility and Alternative Possibilities*, ed. by David Widerker et Michael McKenna (2003). L'argument se trouve p. 143-145.

¹⁴ *Ibid.*, p. 143. Les observations du cerveau en activité suggèrent en effet qu'à chaque acte mental correspond l'activité de toute une zone du cerveau. Par exemple, avec l'IRM (imagerie par résonance magnétique), on peut voir qu'une activité d'imagination s'accompagne d'un flux sanguin important dans une certaine partie du cerveau. On suppose donc que toute cette partie est active pendant l'effort d'imagination.

¹⁵ Stump a probablement en tête le dualisme de Thomas d'Aquin tel qu'elle le comprend. cf. E. Stump, "Non-Cartesian Dualism and Materialism without Reductionnism", *Faith and Philosophy* 12 (1995), p. 505-531. Pour elle, il est possible d'envisager un dualisme dans lequel chaque événement mental est réalisé dans un événement physique simultané.

l'âme n'est pas étendue, et est donc d'une nature profondément différente de celle du corps, ce qui entraîne la difficulté de savoir comment deux substances aussi différentes peuvent agir l'une sur l'autre. Mais rien de tout cela n'est nécessaire dans l'hypothèse dualiste rejetée par Stump. Il suffit de dire que le choix n'est pas un événement physique observable et qu'il cause des événements physiques observables. Si par nature, il ne peut être observé, alors le neuroscope ne permet pas de prévoir le choix d'un individu en observant le début d'une séquence neuronale. Ensuite, Stump suppose qu'un choix est indivisible tout en étant réalisé dans une suite d'événements physiques. Si le choix est effectivement réalisé dans une suite d'événements, on peut penser qu'il s'agit d'un processus. Il pourrait donc y avoir des demi-choix ou des choix avortés. Ainsi, même si Jones ne peut pas parvenir au terme du processus x-y-z, il peut *tendre volontairement* vers ce terme. Jones aurait donc la possibilité de vouloir autrement, même s'il n'a pas la possibilité de choisir autrement.

2° - Il y a un deuxième présupposé critiquable dans l'argumentation de Stump. Elle admet que l'existence d'un événement indéterminé à l'origine de l'action est ce qui la rend libre. Cette thèse est ce qu'on appelle « l'incompatibilisme-source » (*source incompatibilism*). Contrairement à l'incompatibilisme classique, il ne considère pas la possibilité d'agir autrement comme une condition nécessaire de la responsabilité. Pour savoir si un acte est accompli librement, il faut remonter à sa source et voir si l'on peut y trouver un événement indéterminé. Le but de l'exemple imaginé par Stump est de montrer que c'est la véritable condition de la liberté et de la responsabilité morale. Le lecteur doit sentir intuitivement que Jones est responsable de son vote même s'il n'avait pas la possibilité de choisir autrement. Le problème est que certains lecteurs n'ont pas cette intuition, car les possibilités alternatives a/x ne sont pas contrôlées par l'agent. Stump le dit d'ailleurs elle-même : elles ne sont pas « accessibles à l'agent ». En conséquence, l'agent n'est pas responsable du fait que a ou x se produise. Cela survient indépendamment de sa volonté. Plusieurs arguments classiques, que nous examinerons plus loin (en particulier le *Mind argument*) sont construits sur la base de cette intuition : si une action est le résultat d'un événement physique indéterminé, elle est un effet du hasard, et l'agent n'en est pas du tout responsable. Si cette intuition opère dans le *Mind argument*, elle peut également opérer ici, dans l'argument de Stump, ou plutôt contre lui. Nous avons affaire ici à un conflit d'intuitions. Beaucoup de philosophes pensent que si l'action de Jones n'est pas le résultat d'un événement indéterminé *contrôlé par l'agent*,

l'indéterminisme physique n'ajoute rien du point de vue de la responsabilité morale. De même, nous pourrions penser que si le choix, la décision, ou la volition de Jones n'est pas le résultat d'un processus physique contrôlé par lui, il n'en est pas responsable. Imaginons que le choix de Jones soit l'effet d'un processus physique indéterminé survenant dans son petit orteil gauche. Jones ne serait pas responsable de son action car elle découlerait d'un processus aléatoire échappant à son contrôle. Maintenant, qu'en est-il si l'événement indéterminé, au lieu de se produire dans un orteil de l'agent, apparaît dans son cerveau ? La différence est-elle significative ? C'est un point très délicat, mais si nous pensons, comme Stump, que l'événement indéterminé n'est pas contrôlé par l'agent, pourquoi en serait-il responsable ? Et s'il n'en n'est pas responsable, pourquoi serait-il responsable de ce qui en découle de façon inévitable ?¹⁶

Ces remarques montrent que l'intuition invoquée dans l'argument de Stump est liée au présupposé selon lequel si le choix de l'agent résulte d'un processus physique indéterminé, il en est responsable¹⁷. Si le lecteur n'est pas enclin à faire cette hypothèse, il n'a pas l'impression que Jones soit responsable, sauf s'il a mal compris l'exemple, croyant que les possibilités alternatives a/c étaient au pouvoir de l'agent, mais dans ce cas, il réagit en disant qu'il y a ici un frémissement de liberté même s'il n'implique pas la possibilité de *choisir* autrement. En somme, nous pouvons répondre à l'objection de Stump en exposant un dilemme. Dans l'exemple du neuroscope, il y a deux événements possibles à la source du choix : a, et x.

- soit ces possibilités alternatives sont accessibles à l'agent, et on peut y voir des actions de sa volonté, même si ce ne sont pas encore des choix ou des décisions. L'agent a ainsi des possibilités alternatives, et c'est pourquoi il est responsable.
- Soit ces possibilités alternatives ne sont pas accessibles à l'agent. Dans ce cas, il n'est pas responsable de son choix.

¹⁶ J'utilise ici un argument appelé le principe du transfert de non-responsabilité : si A n'est pas responsable de x, et si y est une conséquence nécessaire de x en vertu des lois de la nature, alors A n'est pas responsable de y. Le principe du transfert sous ses différentes formes sera discuté dans le chapitre II.

¹⁷ D'où la référence à Kane, p. 309-310. Kane est le libertarien qui est allé le plus loin dans l'utilisation du concept d'indéterminisme quantique.

Dans un cas comme dans l'autre, l'exemple échoue, car il ne montre pas un agent à la fois responsable de son choix et incapable de vouloir autrement. Pour accepter l'argument de Stump, il faut croire qu'un choix résultant d'un processus indéterministe peut être un choix libre même si l'agent ne contrôle pas ce processus.

2.4 L'argument du dilemme

Dans un article¹⁸ de 1995, David Widerker a présenté un argument permettant de répondre à Stump, mais dont la portée est très générale, car il vise tous les exemples à la Frankfurt, entendons par là les exemples où un agent est responsable de ce qu'il fait mais où quelque chose ou quelqu'un l'empêche d'agir autrement en intervenant si besoin, après avoir détecté un indice de ce que l'agent allait faire. Après cet argument, je montrerai comment les compatibilistes peuvent y répondre, en imaginant des exemples où il n'y a pas d'intervenant contrefactuel.

Selon Frankfurt, la réfutation de PAP suppose d'établir la proposition suivante :

(IRR) Il peut y avoir des circonstances dans lesquelles l'action d'une personne est rendue nécessaire par ces circonstances sans être causée par elles¹⁹.

Dans l'exemple de Frankfurt, la présence du neurochirurgien et son intention de contrôler Jones au moyen d'appareils parfaitement efficaces rend nécessaire l'action de Jones. Il ne peut pas voter pour l'autre candidat. En ce sens, nous pouvons dire que l'action de Jones est déterminée par certaines conditions causales²⁰. Mais comme je l'ai montré plus haut, *déterminer causalement* ne revient pas nécessairement à *causer de manière déterministe*. Dans un exemple à la Frankfurt, l'intervenant contrefactuel ne *cause* pas le résultat. L'agent agit par lui-même et semble donc responsable de son action. Mais précisément, l'objection de

¹⁸ David Widerker, « Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities » (1995).

¹⁹ Frankfurt, "Alternate Possibilities and Moral Responsibility" (1969), p. 830.

²⁰ Du moins si nous considérons l'action de Jones comme un universel pouvant être instancié de plusieurs manières.

Widerker permet de montrer que dans tout exemple à la Frankfurt, l'action est nécessairement déterminée par des causes. En effet, si ce n'était pas le cas, l'intervenant contrefactuel ne pourrait pas prévoir infailliblement ce que va faire l'agent ; et par conséquent, il ne pourrait pas intervenir à temps pour l'empêcher d'agir selon sa propre tendance. Frankfurt aurait donc échoué à décrire une situation dans laquelle le choix de l'agent est rendu nécessaire sans être causé.

Pour discuter l'exemple de Frankfurt, précisons-le par des indications temporelles. Supposons que Jones hésite à tuer Smith (à t3). A son insu, Black veut qu'il décide à t2 de tuer Smith et qu'il accomplisse effectivement cette action. Mais Black veut rester discret et éviter d'intervenir sans nécessité. Or il peut être sûr des faits suivants :

- (1) Si Jones rougit à t1, alors si personne n'intervient, Jones décidera à t2 de tuer Smith.
- (2) Si Jones ne rougit pas à t1, alors si personne n'intervient, il ne décidera pas (à t2) de tuer Smith.
- (3) Or si Black remarque en Jones un signe qu'il ne va pas décider de tuer Smith, c'est-à-dire s'il voit que Jones ne rougit pas à t1, il force Jones à décider (à t2) de tuer Smith; mais s'il voit qu'il rougit à t1, il ne fait rien.

Enfin, supposons que Black n'intervienne pas car :

- (4) Jones rougit à t1, et décide à t2 de tuer Smith pour des raisons personnelles.

En admettant que l'action dont il s'agit soit la décision (de Jones) de tuer Smith, Frankfurt nous donne-t-il ici un exemple convaincant de situation IRR ? Pour cela, il faut que l'agent soit responsable de son action et que celle-ci soit inévitable sans être déterminée par des causes. En effet, si en décrivant la situation, on indique que l'action est déterminée par des causes suffisantes pour qu'elle se produise, cela exclut la responsabilité morale aux yeux d'un libertarien. Le but d'un contre-exemple à la Frankfurt n'est pas de montrer directement que la responsabilité morale est compatible avec le déterminisme, mais de réfuter le principe des possibilités alternatives en montrant un exemple de situation où l'agent est clairement responsable de ce qu'il fait sans avoir la possibilité d'agir autrement. Si l'on réintroduit le déterminisme causal dans l'exemple, la responsabilité de l'agent n'est plus du tout évidente. Or

tel est bien le problème, ici : pour que la décision de Jones soit vraiment inévitable, il faut des causes suffisantes pour qu'elle se produise.

Widerker fonde cette dernière affirmation sur l'argument du dilemme, déjà utilisé par Kane sous une forme moins systématique. Voici l'argument. La proposition (1) ne peut être fondée sur le fait que le rougissement de Jones est (ou manifeste) un événement causalement suffisant (dans les circonstances de l'exemple) pour que Jones décide de tuer Smith. En effet, une telle supposition serait contraire à IRR et un libertarien ne pourrait que la rejeter ou la trouver incompatible avec la responsabilité de Jones. Il faut donc comprendre (1) de la manière suivante :

(1a) Si Jones rougit à t1, alors il va *probablement* décider (à t2) de tuer Smith.

Or (1a) est compatible avec le fait que Jones ne décide pas de tuer Smith. Par exemple, il est capable d'agir en allant contre les tendances de son tempérament. L'autre possibilité est de lire (1) comme un conditionnel de liberté au sens de Plantinga²¹, ce qui donne :

(1b) Si Jones rougit à t1, alors il décidera *librement* à t2 de tuer Smith.

Mais un libertarien peut répondre que dans cette hypothèse, lorsque Jones rougit à t1, il est en son pouvoir de ne pas décider de tuer Smith, même s'il est certain (pour Dieu, par exemple) qu'il décidera à t2 de tuer Smith. Ainsi, dans les deux cas, Jones a le pouvoir de ne pas décider de tuer Smith, donc il n'est pas dans une situation IRR. Le dilemme est le suivant : soit la vérité de (1) est fondée sur un fait causalement suffisant pour que Jones décide à t2 de tuer Smith, soit elle n'est pas fondée sur un tel fait. Dans la première hypothèse, la situation décrite par Frankfurt n'est pas une situation IRR, car le facteur qui rend toute autre décision impossible *produit* la décision. Et dans la deuxième hypothèse, on ne voit pas pourquoi la décision de Jones serait inévitable.

²¹ Cf. Alvin Plantinga, *The Nature of Necessity*, Oxford, Oxford University Press, 1974, ch.9.

Le défenseur des exemples à la Frankfurt peut cependant répondre à l'objection du dilemme en soutenant que l'acte de décision comprend l'inclination et la délibération qui le précèdent. Cela suppose que la décision soit un processus complexe étalé dans le temps. En interrompant ce processus, l'intervenant contrefactuel empêcherait la décision de se produire. Mais il faut alors ajouter que dans ce cas, aucun acte ne se produirait : la délibération et l'inclination ne seraient pas en elles-mêmes des *actions*. Ainsi, la possibilité d'avoir une autre inclination ne serait pas une possibilité alternative, puisqu'elle ne serait pas une possibilité d'agir autrement. Finalement, pour répondre à l'objection du dilemme, le défenseur des exemples à la Frankfurt doit se fonder sur les mêmes présupposés que l'argumentation de Stump. Premièrement, rien n'empêche le libertarien de dire que Jones a des possibilités alternatives, même si elles ont moins de portée que les possibilités auxquelles il pensait au départ. Deuxièmement, il peut croire que la décision est un acte mental simple : le terme de la délibération. Troisièmement, il peut aussi soutenir qu'une inclination ou une délibération interrompue est déjà une action. Il paraît donc difficile d'échapper à l'objection du dilemme, à moins de fonder sa réponse sur une série d'affirmations douteuses.

Un libertarien peut s'accorder avec Frankfurt pour rejeter ce principe :

(PAP') Si A est un acte complexe, une personne est moralement responsable en accomplissant A seulement si elle aurait pu accomplir un autre acte complexe à la place de A.

Mais comme le note Widerker²², Frankfurt n'a donné aucune raison de rejeter les principes suivants :

(PAPD) Une personne est moralement responsable de sa décision de faire A seulement si elle aurait pu décider autrement.

(PAV) Une personne est moralement responsable d'une action A seulement si elle aurait pu éviter de l'accomplir.

²² *Ibid.*, p. 259.

(PAV') Si V est un acte complexe (comme tuer, voter, insulter, etc.), et t^* le moment exact où une personne S fait V intentionnellement, S est moralement responsable de faire V *intentionnellement* à t^* seulement s'il était en son pouvoir de ne pas faire V intentionnellement à t^* .

2.5 Un exemple sans intervenant contrefactuel

Les compatibilistes imaginent des situations à la Frankfurt dans lesquelles l'agent n'a pas la possibilité de vouloir autrement. Dans tous ces exemples, l'intervenant contrefactuel est capable d'anticiper le choix de l'agent en détectant un signe. Par exemple, Gunnar rougit avant de tirer sur Ridley, et blêmit avant de renoncer à tirer. Nous avons vu (en 2.4), que le rôle d'un tel signe est ambigu. Soit il est lié à un fait causalement suffisant pour que l'agent prenne la décision prévue par l'observateur. Dans ce cas, la décision de l'agent est causalement déterminée, et nous avons le droit de conclure qu'il n'est pas libre. Soit le signe indique seulement que l'agent fera probablement ce choix, mais cela est insuffisant pour l'intervenant contrefactuel, car l'agent est capable de déjouer ses prédictions. Il a donc des possibilités alternatives. Par exemple, il rougit, mais au moment où il semble prêt à tirer, il jette son revolver, ou baisse son arme et s'en retourne avec un air méditatif.

Cet « argument du dilemme » a une certaine force. Cependant, il y a plusieurs réponses à cette objection. Widerker lui-même, d'abord incompatibiliste, s'est rallié à la position de Stump en imaginant des situations IRR sans intervenant contrefactuel. Je vais présenter ici cette réponse et la critiquer, après quoi j'examinerai les autres réponses envisageables : en 2.6, celle du blocage parfait, utilisée par Mele, Robb, et Hunt, et en 2.7, celle de McKenna, qui consiste à imposer un critère de « robustesse » des possibilités alternatives.

Widerker a défendu le principe des possibilités alternatives jusqu'à ce qu'il trouvât des contre-exemples échappant à son argument du dilemme. Leur originalité est qu'ils ne supposent pas d'intervenant contrefactuel (ni par conséquent de « signe » permettant de

prévoir le choix de l'agent). Ils ne supposent pas non plus de conditions causalement suffisantes pour que l'agent fasse un certain choix. Voici un tel exemple :

[Personnes-Z] Les personnes-Z sont atteintes d'une étrange maladie : lorsqu'elles sont tentées d'accomplir une action immorale (comme ne pas respecter une promesse), ou bien elles succombent à la tentation, ou bien elles perdent connaissance pendant une seconde puis redeviennent conscientes. Ce phénomène ne répond pas à un déterministe strict. Il est purement aléatoire, c'est-à-dire indéterministe²³.

Imaginons que Jones soit une personne-Z. A t, soit il décide par lui-même de violer sa promesse, soit il s'évanouit. Le seul moyen pour Jones de ne pas violer sa promesse est de s'évanouir. Or cette alternative, précise Widerker, ne lui est pas accessible (« *actionally accessible* »). Il peut seulement s'évanouir, sans avoir choisi de s'évanouir. C'est une possibilité mais qu'il ne peut pas choisir, ni réaliser à volonté. En ce sens, cette possibilité alternative ne lui est pas accessible. Il n'en reste pas moins que dans la situation actuelle, Jones choisit par lui-même de violer sa promesse. Ainsi, Jones paraît bel et bien responsable de sa décision, puisqu'il n'est pas déterminé causalement mais décide par lui-même.

Ceci amène Widerker à corriger sa position. Selon Widerker, la responsabilité morale ne suppose pas la possibilité (accessible à l'agent) de choisir autrement. En revanche, si son choix était le résultat d'un processus entièrement déterministe, il n'en serait pas responsable. Pour que l'agent soit responsable de son choix, il faut donc que celui-ci ne soit pas déterminé causalement. Widerker adopte ainsi la thèse de l'incompatibilisme-source (*source incompatibilism*). Selon cette thèse, la vraie condition métaphysique de la liberté et de la responsabilité morale n'est pas la possibilité de choisir autrement, mais le fait que le choix de l'agent ne soit pas déterminé causalement²⁴.

²³ D. Widerker, "Libertarianism and the Philosophical Significance of Frankfurt Scenarios", *The Journal of Philosophy*, vol. 53, n. 4 (2006), p. 163-187. Cet exemple, que je reformule librement, se trouve p. 169.

²⁴ Widerker distingue « *on his own* » et « *up to him* » : l'action (ou au moins la décision) doit être « *on his own* », c'est-à-dire causée par lui et non par des causes qu'il ne contrôle pas. Elle n'a pas besoin d'être « *up to him* », c'est-à-dire telle que l'agent pouvait ne pas l'accomplir.

L'exemple imaginé par Widerker a plusieurs avantages par rapport à celui de Stump. Comme il ne suppose pas d'intervenant contrefactuel, il échappe à l'objection du dilemme formulée par Widerker quelques années plus tôt. Il paraît aussi plus simple, car il ne repose pas sur des hypothèses de philosophie de l'esprit. Mais si la première impression du lecteur est bien que Jones est responsable de ce qu'il fait, un examen plus attentif de l'exemple proposé par Widerker peut dissiper cette impression. Une question se pose en effet : que veut-on dire en déclarant que Jones choisit *par lui-même* ? Dans l'exemple de Widerker, ce jugement se fonde sur le fait que le processus aboutissant au choix est indéterministe. Le présupposé est donc le suivant :

(AA) Lorsqu'un choix ou une action est l'issue d'un processus indéterministe, celui qui fait ce choix ou cette action le fait *par lui-même*.

Mais la condition énoncée par ce principe (le fait que le choix résulte d'un processus indéterministe) n'est pas suffisante. En effet, si ce processus indéterministe se déroulait à l'extérieur à Jones, dans une machine qu'il ne contrôle pas, le choix de Jones serait finalement *déterminé* causalement par le résultat d'un processus indéterministe échappant entièrement à son contrôle. Le fait que dans l'exemple de Widerker, le processus se déroule à l'intérieur du corps de Jones ne résout pas nécessairement cette difficulté. En effet, nous pourrions imaginer un cas où le processus indéterministe se déroule à l'intérieur de son corps, mais dans une machine greffée par un savant fou. Jones ne choisirait pas plus *par lui-même* que dans le cas précédent. Finalement, on peut supposer qu'il ne choisit vraiment *par lui-même* que s'il contrôle le processus indéterministe aboutissant à son propre choix. Il faut donc savoir à quelles conditions un agent contrôle ses choix.

Nous reviendrons sur ces difficultés au chapitre VII. Mon intention, en faisant ces remarques, n'est pas de trancher la question dès maintenant mais de montrer simplement que l'interprétation de l'exemple proposé par Widerker n'est pas aussi aisée qu'il n'y paraît à première vue. Le fait que le processus aboutissant au choix ne soit pas un processus déterministe ne garantit pas de manière évidente que l'agent fasse ce choix *par lui-même*. Et s'il faut que l'agent contrôle ce processus, il est permis de croire qu'il doit surtout en contrôler l'issue. Cela est-il possible si l'agent n'a pas le pouvoir de faire un autre choix ? A la

réflexion, l'interprétation de l'exemple admise par Widerker repose comme l'argument de Stump sur des hypothèses auxiliaires ou des prémisses cachées. Cela limite beaucoup sa valeur, car plus un exemple est complexe, moins le jugement intuitif que nous portons sur la responsabilité morale de l'agent est fiable. Si l'exemple de Widerker est convaincant à première vue, c'est peut-être parce qu'il reste flou sur la question que je viens de poser. Lorsqu'on réfléchit à cette question, l'intuition que Jones est responsable se dissipe ou perd une grande partie de sa force.

On peut même se demander si le scénario envisagé par Widerker est cohérent. Widerker considère que Jones est responsable de son action. Cela suppose qu'il exerce un certain pouvoir en accomplissant cette action, et qu'il ait accès à cette action, sans avoir accès à l'autre possibilité (s'évanouir). Un exemple à la Widerker doit donc avoir les caractéristiques suivantes :

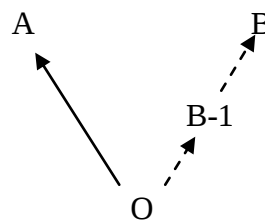
1. S fait A de lui-même et est responsable de A.
2. A est l'issue d'un processus indéterministe dont le résultat peut être aussi que S fasse B.
3. B n'est pas une possibilité accessible à S.

Cette description pose la question suivante : si la possibilité de faire A est accessible à S, pourquoi la possibilité de faire B ne le serait-elle pas aussi ? Notons d'abord que dans l'exemple de Widerker, la possibilité B n'est pas bloquée par un intervenant contrefactuel : Jones *peut* s'évanouir. Deuxièmement, puisque B et A sont les deux issues possibles d'un même processus indéterministe, B *peut* se produire dans des circonstances exactement identiques à celles de l'action A. Pour que Jones fasse B, il n'est pas nécessaire que se réalise une condition ou un événement qui n'existe pas dans la situation actuelle (où Jones fait A). Si donc la possibilité de faire A est accessible à l'agent, on ne voit pas comment la possibilité de faire B ne lui serait pas également accessible.

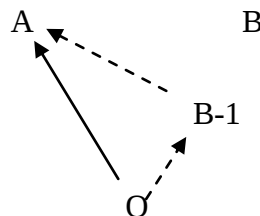
On peut se demander si la force de l'exemple proposé par Widerker ne dépend pas d'une confusion entre différentes sortes de possibilités alternatives. Quand nous décrivons des dispositions d'une substance chimique ou d'un êtres vivants, nous faisons référence à des

possibilités alternatives qui ne leur sont pas accessibles. Par exemple, un morceau de sucre qui est au sec dans la situation actuelle peut se dissoudre dans l'eau, mais il n'a pas accès à cette possibilité, car celle-ci ne s'actualise que dans certaines conditions non réalisées dans la situation actuelle. Mais le cas d'un processus indéterministe est fort différent. Si A et B sont les deux issues possibles d'un même processus indéterministe, alors les conditions de A sont les mêmes que les conditions de B, et il n'y a rien de plus à faire pour obtenir B que pour obtenir A.

Pour éviter cette conclusion, il faut imaginer un processus tel que la réalisation de B (la possibilité alternative) suppose au moins une condition nécessaire que ne suppose pas la réalisation de A. Dans ce cas de figure, l'agent peut avoir accès à A sans avoir accès à B, car la réalisation de B exige plus que la réalisation de A. Par exemple, supposons qu'il ait accès à B-1 autant qu'à A, ces deux événements étant les issues possibles d'un même processus indéterministe. B-1 est une condition nécessaire mais non suffisante de B. Alors, bien que l'agent ait accès à B-1, il n'a pas nécessairement accès à B.



On peut d'ailleurs imaginer qu'un intervenant contrefactuel empêche B-1 d'aboutir à B. Par exemple, si B-1 se produit, il y a un processus dans le cerveau de l'agent qui détecte B-1 et l'empêche d'aboutir à B. Il interrompt la chaîne causale aboutissant normalement à B, et produit A à la place de B.



Mais ce cas de figure tombe sous les objections de Widerker lui-même dans sa présentation de l'argument du dilemme. Celui-ci porte en effet sur tous les exemples où la possibilité alternative est censée être exclue par un intervenant contrefactuel qui détecte une tendance de l'agent. En outre, dans les deux cas que je viens d'esquisser, si l'agent n'a pas accès à B, il a tout de même accès à B-1. Il y a donc une possibilité alternative accessible à l'agent. Pour refuser cette conclusion, on pourrait s'appuyer sur le fait que B n'est pas une action, du moins dans l'exemple de Widerker, où B consiste à s'évanouir. Mais c'est peut-être que l'exemple est incohérent. En effet, si A est une action, pourquoi B (ou B-1) ne le serait-elle pas aussi ? Ce que nous « faisons » n'est pas toujours une action : par exemple, lorsque nous digérons; mais ce qui distingue les actions des autres choses que nous faisons, c'est le contrôle que nous exerçons sur nos actions. Si A est une action de Jones, et s'il exerce le même contrôle sur B (lorsqu'il fait B) que sur A (lorsqu'il fait A), je ne vois pas pourquoi B ne serait pas une action au même titre que A.

2.6 Surdétermination et blocage parfait

Nous avons vu que les exemples à la Frankfurt ne peuvent fonctionner que s'ils prouvent l'affirmation IRR :

(IRR) Il peut y avoir des circonstances dans lesquelles l'action d'une personne est rendue nécessaire par ces circonstances sans être causée par elles.

En effet, si l'action est causée ou produite par les circonstances, elle ne vient pas de l'agent lui-même, et le libertarien en déduira que l'agent n'est pas responsable de son action. L'intérêt des exemples à la Frankfurt est justement que *sans supposer un déterminisme causal*, ils montrent un agent qui n'a pas la possibilité d'agir autrement mais qui paraît responsable de son acte.

Pour éviter toute ambiguïté, il convient de préciser dans l'exemple que l'action et le choix dont elle provient résultent d'un processus indéterministe. C'est ce que font Stump et

Widerker (le « second » Widerker). Or une fois cela dit, il ne peut y avoir que deux espèces de situations IRR :

- soit il y a encore des possibilités alternatives dans l'agent (antérieures à l'action ou même au choix).
- Soit il y a un blocage parfait qui ne laisse subsister aucune possibilité alternative dans l'agent, pas même celle d'incliner vers un autre choix. Les possibilités alternatives ne sont pas dans l'agent lui-même, mais dans un processus physique indéterministe extérieur à l'agent et qui aboutit à son choix.

Les situations IRR imaginées par Stump et Widerker sont du premier type, et nous avons vu qu'elles posent trop de difficultés d'interprétation pour être convaincantes. Alfred Mele et David Robb ont imaginé des situations du second type :

Notre scénario met en scène un agent, Bob, qui vit dans un monde où le déterminisme est faux. Cela ne veut pas dire qu'aucun événement de ce monde n'est causé de façon déterministe. Par exemple, une certaine particule radioactive s'est désintégrée dans la ville de Bob à midi, et cet événement n'était pas déterminé causalement, mais la particule a agi comme détonateur aléatoire pour une bombe, et après sa désintégration, la bombe a été déterminée causalement à exploser.

A t1, Black initie un processus déterministe P dans le cerveau de Bob pour produire en lui la décision de voler la voiture d'Anne (décision qu'il prendra une heure plus tard, à t2). Le processus (...) aboutira à cette décision, sauf si Bob décide par lui-même à t2 de voler la voiture ou s'il est incapable à t2 de prendre une décision (parce qu'il est mort à t2, par exemple). Le processus ne dépend nullement d'un « signe » indiquant ce que Bob va décider. Or, de fait, Bob décide par lui-même de voler la voiture, sur la base de sa propre délibération indéterministe, et sa décision n'est pas déterminée par une cause. Mais si, à ce moment précis, il n'avait pas décidé par lui-même de la voler, P aurait abouti de façon déterministe, à t2, à sa décision de la voler. Néanmoins, il est certain que P n'influence nullement le processus indéterministe aboutissant effectivement à la décision de Bob.²⁵

Bob est responsable de ce qu'il fait. Sa décision n'est pas causée par un processus déterministe, ni par un processus indéterministe extérieur à lui ; elle vient de sa propre activité

²⁵ Alfred Mele et David Robb, « Rescuing Frankfurt-style cases », *Philosophical Review* 107 (1998), p. 97-112, repris dans Fischer (2005a, vol.3). Le passage cité est p. 333 dans l'anthologie de Fischer.

indéterministe de délibération. Ainsi, Bob semble être responsable alors qu'il n'a pas la possibilité de prendre une autre décision, le processus P venant canaliser l'agent juste au moment où il prend sa décision. Mais on peut douter qu'un tel scénario soit cohérent. En effet, comment Bob peut-il être responsable de sa décision si au moment du choix, sa volonté est bloquée par un processus déterministe ?

Pour montrer que l'exemple est cohérent, Mele et Robb imaginent une machine conçue pour produire des effets artistiques de différentes formes et couleurs. Chaque couleur obtenue est déterminée par celle d'une balle (bb) frappant le récepteur de la machine à un certain moment. Des pistolets à peinture projettent la peinture contenue dans les balles.

Si une balle de couleur x frappe le récepteur de la machine M, et si M n'est pas déjà en train de fabriquer une « œuvre », cela enclenche le processus de fabrication qui aboutit à un effet de couleur x . Mais la machine a été conçue de telle sorte que si deux balles (ou plus) frappent en même temps le récepteur, M réagit en produisant un objet de la même couleur que la balle la plus à droite.

Ce qui posait problème dans l'exemple de Bob était le fait que le choix libre et le processus déterministe coïncidaient. Cette situation est comparable à ce qui se passe dans la machine si deux balles de même couleur (par exemple, bleues) frappent en même temps le récepteur. La balle de droite (qui représente le processus par lequel Bob décide librement) arrive en même temps que la balle de gauche, également bleue. La balle qui vient par la gauche ne *cause* pas la projection de peinture bleue. De la même façon, le processus P ne cause pas la décision de Bob.

Cela dit, on peut toujours se demander ce qui arrive lorsque les deux processus divergent. Voici comment Mele et Robb traitent cette difficulté :

Le problème peut se représenter avec fantaisie de la manière suivante. Deux « nœuds de décision » différents sont directement concernés dans le cerveau de Bob. L'allumage du nœud N1 représente sa décision de voler la voiture, et l'allumage du nœud N2 représente sa décision de *ne pas* voler la voiture. Dans des circonstances normales et en l'absence de préemption, lorsqu'un processus « frappe » en Bob un nœud de décision, il allume ce nœud. S'il arrivait à la fois que P frappe N1 à t_2 et que x ne frappe pas N1 à t_2 , alors P allumerait N1. Si les deux processus à la fois frappaient N1 à t_2 , ce ne serait pas P, mais la

délibération de Bob, processus indéterministe, qui allumerait N1. La question qui se pose est alors celle-ci : qu'arriverait-il si à t_2 , P frappait N1 et x , N2 ? Autrement dit, qu'arriverait-il si les deux processus « divergeaient » ainsi ? Et pourquoi ?

Nous pouvons prolonger l'histoire de Bob de la manière suivante. Si les deux processus frappaient N1 à t_2 , la délibération de Bob (donc le processus indéterministe x) devancerait P et allumerait N1, mais il est également vrai que si à t_2 , P frappait N1 et x , N2, P l'emporterait. Dans le second cas, P allumerait N1 et le processus indéterministe n'allumerait pas N2. Bien sûr, les lecteurs attendent une histoire montrant pourquoi, bien que x devance P dans la première situation, P l'emporte sur x dans la seconde. En voici une. A t_2 , P a « neutralisé » N2 (mais sans influencer ce qui arrive en x). C'est pourquoi si x frappait N2 à t_2 , N2 ne s'allumerait pas. Plus précisément, à t_2 , P a neutralisé en Bob toutes les nœuds permettant des décisions incompatibles avec celle de voler la voiture d'Anne (par exemple, une décision à t_2 de ne voler aucune voiture, et une décision à t_2 de ne rien voler du tout). Pour résumer cela de façon commode, disons qu'à t_2 , P a neutralisé N2 et toutes les nœuds de décision du même genre que N2. Rappelons-nous que nous n'avons besoin que d'un scénario conceptuellement possible, et celui-là a certainement l'air d'en être un.²⁶

Cette situation réunit donc les caractéristiques suivantes, qui seraient parfaitement compatibles :

- 1) Dans tout futur possible ouvert à Bob après le commencement de P, et où il est capable de prendre une décision (nous appellerons cela un futur du genre F), Bob décide à t_2 de voler la voiture.
- 2) Dans tout futur F, si un processus indéterministe de prise de décision (par Bob) -un processus du genre X – n'aboutit pas à la décision (de Bob) de voler la voiture, P produit cette décision de façon déterministe.
- 3) Dans tout futur F, si un processus du genre X aboutit à la décision (de Bob) de voler la voiture, P ne produit pas cette décision.
- 4) Un processus du genre X aboutit à la décision (de Bob) de voler la voiture, et P ne produit donc pas cette décision.

²⁶ *Ibid.*, p. 104, ou p. 335 dans l'anthologie de Fischer (2005a, vol.3).

Selon Mele et Robb, ces caractéristiques sont cohérentes entre elles. La situation qu'ils présentent est conceptuellement possible. Elle prouve qu'une décision peut être inévitable sans être déterminée causalement, et qu'on peut être responsable d'une décision et de ce qui en découle sans avoir eu la possibilité de ne pas prendre cette décision. L'exemple semble être une réponse satisfaisante à l'objection du dilemme, car il ne suppose pas d'intervenant contrefactuel utilisant un signe pour contrôler l'agent. Il pose néanmoins quelques difficultés.

Tout d'abord, dans cet exemple, l'intervenant contrefactuel est remplacé par un processus déterministe qui, à la différence du premier, agit effectivement dans le monde actuel. Dans les exemples à la Frankfurt, ce qui permet de dire que l'agent est responsable est le fait que l'intervenant contrefactuel n'intervient pas dans le monde actuel. Or il n'en est pas de même ici : P agit effectivement, et ce faisant, détermine le choix de Bob. Autrement dit, dans les exemples à la Frankfurt, c'est la simple présence d'un certain facteur (l'intervenant contrefactuel) qui détermine, dans les conditions spécifiées dans l'exemple, le choix de l'agent ; alors qu'ici, le facteur P n'est pas seulement présent mais détermine *par son action* le choix de Bob. C'est pourquoi on peut avoir l'impression que l'exemple de Mele et Robb est un cas de « déterminisme pur et simple », comme le dit Kane, même si avant t₂ (l'instant du choix), le processus amenant la décision a des aspects indéterministes. Kane, Ekström et Pereboom²⁷ ont le sentiment que le blocage détermine causalement la décision de Bob. Une fois que le processus P est enclenché, il garantit le fait que Bob choisira de voler la voiture. La responsabilité ultime du vol semble donc appartenir à l'individu qui enclenche P plutôt qu'à Bob lui-même.

Mele et Robb répondent à cette objection en distinguant ce qui est *inévitabile*, et ce qui est *causé de manière déterministe*. Les exemples classiques à la Frankfurt montrent qu'on peut

²⁷ Robert Kane a formulé cette réponse dans une discussion avec Fischer au symposium intitulé « The Significance of Free Will in Philosophy and Phenomenological Research 60 » (2000), p.162, puis dans « Responsibility, Indeterminism and Frankfurt-style Cases : A Reply to Mele and Robb », in Widerker & McKenna (2006), p. 97. Pereboom aussi défend cette réponse dans *Living without free will* (2001), p. 17-18. Ekström répond à Mele et Robb dans « Libertarianism and Frankfurt-style cases », p. 313-315 dans Kane (2002b).

rendre inévitable x sans *causer* x , et donc sans le causer de manière déterministe²⁸. Le fait que P rende inévitable la décision de Bob ne signifie donc pas que P la détermine causalement. Néanmoins, le principe suivant est plausible :

Si P rend nécessaire x en agissant dans le monde actuel, P détermine x causalement.

Ici, P (avec les circonstances de t_2 et les lois de la nature) rend nécessaire la décision de Bob en agissant dans le monde actuel. Cela semble suffisant pour que P détermine causalement le choix de Bob. Agir peut consister à installer un dispositif qui n'agit pas dans le monde actuel mais qui est programmé pour empêcher Bob de ne pas faire le choix de voler la voiture. Mais cette action ne permet pas un blocage parfait, puisqu'il laisse à Bob des possibilités alternatives dans ses inclinations. Pour que le blocage soit parfait, il faut que l'action de P atteigne x , c'est-à-dire qu'il y ait une chaîne causale allant de P à x , ou si l'on préfère, que l'influence déterminante de P s'étende jusqu'au choix de Bob. Mele et Robb reconnaissent que P est suffisant *nomologiquement* et *contrefactuellement* pour que Bob décide à t_2 de voler la voiture. Mais cela, pensent-ils, ne revient pas à dire que P *cause* cette décision. C'est bien x , un processus indéterministe, qui cause la décision. Il en serait ici comme dans l'exemple utilisé par Locke : un homme se réveille dans une chambre fermée à clef de l'extérieur, mais en compagnie d'une personne qu'il aime. Le fait qu'il reste dans la chambre n'est pas causé par le blocage (le fait qu'on ait fermé la porte à clef). Celui qui a verrouillé la porte n'a pas causé le fait que l'homme endormi demeure dans la chambre, pourtant il a rendu ce fait nécessaire.

Mais un autre problème se pose : selon Widerker et O'Connor²⁹, x ne peut empêcher P de produire la décision à t_2 . Pour cela, il faudrait que x intervienne *avant* t_2 . Or dans l'exemple, la décision ne peut se produire avant t_2 . L'exemple est donc incohérent, et si on le modifie en disant que la décision peut être produite par x avant t_2 , il n'y a plus de blocage, car Bob peut décider de ne pas voler la voiture.

²⁸ Cf. 1.8, p. 44, où j'ai expliqué cette distinction.

²⁹ Cf. David Widerker, « Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look », *Philosophical Perspectives* 14 (2000), p. 181-201, repris et corrigé dans « Blameworthiness and Frankfurt's Argument Against the Principle of Alternative Possibilities », p.53-73 in Widerker & McKenna (2006).

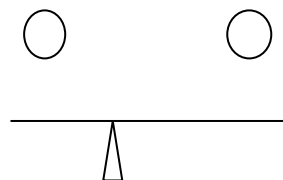
Carl Ginet³⁰ fait à peu près la même objection sous une forme plus détaillée. Dans l'exemple,

- x , en causant la décision, empêche P de la causer à t_2 . Mais
- Si x n'avait pas causé la décision à t_2 , P l'aurait causée à ce même instant.

Selon Ginet, ces deux suppositions sont incompatibles. Comme x est un processus indéterministe, il doit y avoir des possibilités alternatives ouvertes à l'agent jusqu'à t_2 inclus, le moment du choix libre. Mais pour que P rende nécessaire la décision (à t_2) de voler la voiture, il faut que Bob ne puisse pas devancer le processus P jusqu'à t_2 inclus ; donc il faut qu'il n'ait pas la possibilité de choisir autrement jusqu'à t_2 inclus, ce qui contredit la supposition précédente.

Selon Mele et Robb, cette objection n'est pas décisive, car x pourrait empêcher P de causer le choix sans intervenir avant t_2 . Dans les exemples de préemption les plus familiers, on empêche un facteur d'agir en intervenant avant lui. Mais il est possible, selon Mele et Robb, de concevoir une « préemption simultanée » (« *occurrent preemption* »). Par exemple, dans la machine artistique décrite plus haut, la boule de couleur x peut empêcher une autre boule d'entrer dans le pistolet à peinture en arrivant au même endroit et en même temps qu'elle.

C'est le cas si les deux boules tombent simultanément sur une sorte de balancier qui se présente ainsi :



³⁰ C. Ginet, « In Defense of the Principle of Alternative Possibilities » (2000), p. 85-86 dans Widerker & McKenna (2006).

Supposons que toutes les boules pouvant arriver à droite soient de couleur x . Si les deux boules sont de même masse, le balancier penche à droite après l'impact des deux boules à t_2 (l'instant du choix). Ensuite, on imagine aisément la boule de droite tombant dans un entonnoir, puis dans le pistolet à peinture.

Soit P le processus déterministe qui amène la boule bleue sur le balancier M à t_2 , et x le processus indéterministe qui amène la boule rouge sur le balancier M au même instant. L'exemple est cohérent mais on peut se demander si le choix de voler la voiture peut être figuré correctement par l'événement « la boule rouge tombe en M à t_2 ». Ne faut-il pas distinguer deux temps : celui où la boule rouge arrive en M , et celui où le balancier penche vers la droite et libère la boule rouge ? On peut avoir le sentiment qu'en arrivant en M , la boule rouge bloque le processus P , donc l'empêche d'arriver à son terme, alors que P , dans l'exemple, est censé aboutir à son terme de façon inévitable.

Selon Mele et Robb, le choix se produit au moment de l'engagement (« *commitment* »), c'est-à-dire à t_2 , moment où les deux boules arrivent en M , la suite des événements en découlant de façon nécessaire. Or par hypothèse, le choix est le même si P le produit ou si x le produit. Dans l'exemple du balancier, le choix ne peut donc pas correspondre à l'événement « une boule rouge frappe M à t_2 », ni à « une boule bleue frappe M à t_2 », ni encore à « une boule rouge et une boule bleue frappent M à t_2 », car aucun de ces événements ne peut être produit par P tout seul et par x tout seul. Le choix devrait plutôt correspondre à un événement postérieur, comme « la boule rouge entre dans l'entonnoir » (ou dans le pistolet à peinture). Or cela implique que l'exemple de Mele et Robb n'est pas un exemple de préemption simultanée. En effet, si t_2 est l'instant du choix, le moment où les deux boules tombent en M est antérieur. Ginet semble donc avoir raison : x ne peut empêcher P de causer le choix qu'en intervenant avant t_2 . Ainsi, l'exemple utilisé par Mele et Robb n'est pas assez convaincant.

David Hunt a présenté un autre exemple de blocage plus simple, dans lequel, à la différence de l'exemple que nous venons de voir, il n'y a qu'un seul scénario possible. Cette différence met en évidence un aspect du problème : le lien entre l'indéterminisme et la responsabilité morale. « AC » (« autres conditions ») désigne l'ensemble des conditions de la

responsabilité morale autres que PAP, donc les conditions admises aussi bien par les compatibilistes que par les incompatibilistes (conditions épistémiques par exemple).

Imaginons [...] un mécanisme qui bloque non plus les portes mais les entrées des neurones (*neural pathways*). Supposons que la série actuelle des états mentaux de Jones conduisant au meurtre de Smith soit compatible avec AC³¹, mais que le mécanisme soit en train de fonctionner. Le mécanisme n'intervient pas directement dans la série elle-même ; il la laisse se produire d'elle-même, et bloque seulement toutes les alternatives à cette série. Bien sûr, il ne peut pas bloquer les alternatives *en réponse* à la manière dont la série se produit, puisqu'alors, le blocage arriverait trop tard pour avoir un effet sur l'inévitabilité des actions de Jones.

C'est donc plutôt à l'avance que le mécanisme bloque les alternatives, mais par une fantastique coïncidence, les entrées qu'il bloque sont précisément celles qui ne seront nullement empruntées, alors que la seule voie qui demeure ouverte est précisément la route que les pensées de l'individu suivront à coup sûr (si les autres entrées neuronales sont débloquées). Dans ces conditions, l'individu semble encore responsable de ses pensées et actions [...] ». ³²

Contrairement à l'exemple de Mele et Robb, celui de Hunt suppose une coïncidence, comme il le dit lui-même. C'est peut-être un inconvénient. En revanche, si l'argument de Hunt est recevable, il va plus loin, car il montre que la responsabilité morale ne suppose pas l'existence de plusieurs scénarios possibles. Dans l'exemple de Mele et Robb, il y avait plusieurs scénarios possibles avant l'instant du choix. Ici, la série d'événements mentaux ne peut jamais prendre une autre voie que dans le monde actuel. Pour cette raison, Pereboom estime que l'exemple est trop ambigu pour être convaincant. En effet, on peut se demander s'il ne s'agit pas d'un processus déterministe. D'un autre côté, comme dans l'exemple de Mele, on ne peut pas dire que le processus de blocage *cause* la décision. Il la rend nécessaire, mais sans la causer. Sa cause est le processus qui entraîne la série d'événements mentaux aboutissant à la décision. Comme le dit Fischer³³, il est essentiel, pour bien interpréter cet exemple, de savoir si le signal qui aboutit à la décision peut dévier légèrement de sa trajectoire en heurtant les obstacles qui l'empêchent d'aboutir à un autre choix. La simple pression exercée à

³¹ D. Hunt utilise en fait l'abréviation PĀP. Je la remplace par AC pour éviter toute confusion avec PAP.

³² David Hunt, « Moral Responsibility and Unavoidable Action » (2000), p. 375-376 dans l'anthologie de Fischer (2005a).

³³ J. M. Fischer, « Frankfurt-type Examples and Semi-Compatibilism », in Kane (2002a), p. 296.

l'intérieur du « chemin neuronal » contre les parois ou obstacles qui canalisent le signal nerveux et excluent les autres décisions pourrait être interprété comme un effort pour éviter la décision. La possibilité d'exercer cet effort est déjà une possibilité alternative ou un « flicker of freedom ». Or si le signal mental est tellement canalisé qu'il n'y a même plus, à aucun moment, cette possibilité alternative, le processus qui entraîne la décision semble être un processus purement déterministe. Mais alors, l'agent est-il responsable de sa décision ?

Tel qu'il est présenté par Hunt, l'exemple n'est pas suffisamment précis pour mettre en évidence une intuition compatibiliste. L'impression du lecteur, comme Pereboom et Fischer, est que la situation décrite par Hunt est ambiguë parce que la description ne mentionne pas tous les faits qui pourraient compter dans l'interprétation de l'exemple. L'impression que Jones est responsable de son choix peut venir du fait que même si la série d'événements mentaux est parfaitement canalisée, Jones avait la possibilité de ne pas entraîner cette série. Si le fait qu'une impulsion électrique soit envoyée dans ce canal dépend de lui, au sens où il a la possibilité de ne pas donner cette impulsion, il paraît bien responsable de son choix malgré le processus de blocage. Mais pour que l'exemple de Hunt soutienne les intuitions compatibilistes, il faut que Jones ne contrôle ni la trajectoire du signal, ni le fait qu'il soit donné à t_1 (ni même sa vitesse).

Soit i l'impulsion électrique qui cause la décision de tuer Smith, B le processus de blocage, et C la décision de tuer Smith. La situation serait celle-ci :

- Jones ne peut pas empêcher que i se produise à t_1 .
- Jones ne peut pas empêcher que B se produise à partir de t_1 .
- Jones ne peut pas empêcher que (si i se produit à t_1 , et si B se produit à partir de t_1 , C se produit à t_2).

Et tous ces événements se produisent ainsi, donc i se produit à t_1 , puis B , puis C à t_2 .

Pour qu'il n'y ait aucune possibilité alternative dans l'exemple, il faut qu'il corresponde à cette description. Mais s'il tient compte des quatre faits énumérés ci-dessus, un libertarien ne

peut pas se laisser convaincre par l'exemple de Hunt. Celui-ci ne prouve donc pas qu'on puisse être responsable d'une décision sans avoir la possibilité de ne pas la prendre³⁴.

Dans l'exemple de Mele, l'impression que Jones est responsable ne tient pas à une lacune dans la description de l'exemple, mais au fait que la décision de voler la voiture dérive d'un processus indéterministe. Ce caractère indéterministe donne l'impression que Jones a bien pris la décision par lui-même et qu'il contrôlait sa délibération jusqu'à l'instant du choix. Mais comme j'ai essayé de le montrer après Kane et Ginet, le choix lui-même lui échappe, et puisqu'il ne contrôle pas son choix, il ne contrôle pas non plus la délibération qui précède et dont le seul but est d'aboutir au bon choix.

Ainsi, aucune stratégie de blocage n'est convaincante pour soutenir la thèse compatibiliste. Le blocage implique que la décision soit causée de façon déterministe, par préemption, ou par un blocage parfait qui canalise entièrement le signal nerveux de l'agent. Contrairement au blocage parfait imaginé par Hunt, la « préemption simultanée » envisagée par Mele et Robb laisse subsister des possibilités alternatives avant le choix, mais celles-ci ne comptent plus au moment du choix, et la possibilité même d'une préemption simultanée est douteuse, comme nous l'avons vu. Puisque tous les exemples de blocage appartiennent nécessairement à une de ces catégories, il faut conclure que ce genre d'exemple ne peut soutenir de façon convaincante la position compatibiliste. Finalement, il faut revenir à des exemples où il y a des possibilités alternatives.

³⁴ Une autre objection est que l'exemple suppose une coïncidence extraordinaire. Il n'exclut donc pas qu'il y ait un lien essentiel entre la responsabilité et les possibilités alternatives. Un lien conceptuel entre une notion A et une notion B suppose que A ne puisse pas aller sans B, ou l'inverse. Cette impossibilité doit-elle être comprise comme une impossibilité absolue, ou comme une improbabilité extrême ? On pourrait penser avec Aristote, que les propriétés appartenant à l'essence de l'homme sont celles que presque tous les hommes possèdent. En tout cas, l'exemple de Hunt n'exclut pas que l'existence de la responsabilité morale suppose dans presque tous les cas l'existence de possibilités alternatives. Autrement dit, même s'il était convaincant, il ne prouverait pas que la responsabilité morale puisse exister *autant que nous le pensons ou aussi souvent que nous le pensons* dans l'hypothèse où le déterminisme est vrai. Tout ce qu'il peut établir est la compatibilité entre le déterminisme causal et de très rares actions libres.

2.7 Valeur et portée du vacillement de liberté. La « robustesse » des possibilités alternatives

Puisque les stratégies de blocage ne sont pas convaincantes, il semble inévitable que la meilleure stratégie compatibiliste s'appuie sur des exemples où il y a des possibilités alternatives, car là où il n'y a pas de blocage, il y a des possibilités alternatives, aussi minimes soient-elles. Or nous avons vu que si ces possibilités alternatives ne sont pas *accessibles à l'agent*, un libertarien peut refuser l'exemple à volonté en déclarant que l'agent n'est pas responsable de ce qu'il fait. La seule possibilité qui reste, pour défendre la thèse compatibiliste, est donc utiliser un exemple où il y a des possibilités alternatives *accessibles à l'agent mais non significatives du point de la responsabilité morale*.

Avant d'examiner cette stratégie, essayons de voir s'il y a une catégorie de possibilités alternatives qui existe nécessairement dans tous les exemples à la Frankfurt. Selon William Rowe³⁵, qui s'inspire lui-même de Thomas Reid³⁶, un individu n'est responsable d'une action ou d'un choix que s'il a la possibilité de ne pas le causer-en-tant-qu'agent (*to agent-cause*). Même s'il ne peut pas éviter de décider de tuer Smith, il peut éviter de causer *par lui-même* cette décision (dans ce cas, elle est provoquée artificiellement par l'intervenant contrefactuel).

Selon Keith Wyma³⁷, l'agent est responsable de *x* s'il a la possibilité de ne pas engager sa responsabilité dans *x* (son choix ou son action). Selon Michael Otsuka³⁸, l'agent doit avoir

³⁵ Cf. William Rowe, "Two Concepts of Freedom". *The Proceedings and Adresses of the American Philosophical Association* 61 (1987), p. 43-64.

³⁶ Thomas Reid, *Essays on The Active Powers of the Human Minds: An Inquiry into the Human Mind on the Principle of Common Sense* (1788).

³⁷ Plus précisément, Wyma reformule ainsi le principe des possibilités alternatives :

« PPPB (Principle of Possibly Passing the Buck) :

Une personne est moralement responsable de A (quelque chose qu'elle fait) seulement si elle a manqué de faire B quelque chose qu'elle aurait pu faire et qui est tel que faire B l'aurait rendue non-responsable de A.

PPPB requiert un pouvoir contrefactuel sur le fait d'être responsable de A. C'est pourquoi il peut s'énoncer ainsi :

Une personne est moralement responsable de l'action A seulement si, pour quelque action B, (i) elle aurait pu faire B (mais ne l'a pas fait), et (ii) si elle avait fait B, elle n'aurait pas été responsable de l'action A ». (Keith Wyma, « Moral Responsibility and Leeway for Action », *American Philosophical Quarterly* 34 (1997): 57-70, repris dans Fischer (2005a). Ce passage se trouve au vol. 3, p.317, dans l'anthologie de Fischer).

³⁸ Michael Otsuka, "Incompatibilism and the Avoidability of Blame", *Ethics* 108, p. 685-701.

la possibilité de ne pas être blâmable. Toutes ces formulations expriment au fond la même idée si on admet que la blâmabilité suppose la responsabilité morale, et que la responsabilité morale suppose la capacité décrite par Thomas Reid : celle de « causer » ou d'être la source ultime de ses actions en ayant le pouvoir de ne pas les causer (ou de ne pas en être la source ultime). Or ces critères résistent à tous les exemples à la Frankfurt où il y a des possibilités alternatives accessibles à l'agent, mais où celui-ci ne peut pas agir ou même choisir autrement. On peut le constater ici :

[Neuroscope*] Comme dans l'exemple de Stump, Grey est un neurochirurgien surdoué. Il a installé dans le cerveau de Jones un appareil permettant d'observer tous ses états neuronaux et de contrôler ses choix. Grey sait que si Jones veut voter pour les candidats républicains, cet acte de volonté est corrélé à l'activation des neurones a, b et c, et que s'il choisit de voter pour les candidats démocrates, ce choix est corrélé à l'activation des neurones x, y, z. Grey veut éviter que son client vote pour le parti démocrate. Donc s'il observe l'activation de x ou de x et y, il déclenche la séquence a-b-c avant que z ne soit activé. Or Jones choisit par lui-même de voter pour le parti républicain, et Grey n'a donc pas besoin d'intervenir.

Supposons que dans cet exemple, contrairement à celui de Stump, le fait que a soit activé (et non x) ne relève pas d'un processus physique indéterministe échappant au contrôle de Jones. Dans ce cas, Jones paraît responsable de son choix. Or même s'il n'a pas la possibilité de choisir le parti démocrate, il a le pouvoir d'incliner sa volonté vers le vote démocrate. S'il fait cela, il évite de causer-en-tant-qu'agent la séquence a-b-c et son vote républicain, car c'est le neurochirurgien qui déclenche cette suite d'événements. En même temps, il évite d'engager sa responsabilité dans cette suite d'événements, et ainsi d'être blâmable. Les critères de Rowe, Wyma, et Otsuka, sont donc corrects même dans les exemples où les possibilités alternatives accessibles à l'agent se réduisent au minimum. Or nous avons vu dans les deux sections précédentes qu'il ne paraît pas possible d'inventer des exemples où un individu est responsable sans avoir le moindre « *flicker of freedom* ».

Ainsi, la dernière solution qui reste encore, pour réfuter le principe des possibilités alternatives, est de montrer que ce *flicker of freedom* est négligeable du point de vue de la

responsabilité morale. Selon McKenna, deux conditions doivent être réunies pour qu'une possibilité alternative mérite d'être prise en compte :

1° « L'alternative doit être *moralelement significative*. Elle doit nous dire quelque chose (d'autre que ce que nous constatons dans le monde actuel) sur la qualité morale du comportement de l'agent lorsqu'il agit selon ce scénario alternatif.

2° L'alternative doit être *accessible à l'agent*. »³⁹

Stump et Widerker, dans les exemples que j'ai discutés plus haut, se réfèrent à la deuxième condition. Stump imagine un appareil permettant de contrôler les choix de l'agent en interrompant les séquences neuronales qui ne l'orientent pas vers le choix désiré par l'intervenant contrefactuel. Face à cet exemple, la première réponse du libertarien est que l'agent a la possibilité de ne pas faire ce choix *de lui-même* (ce qui correspond à l'idée de Rowe). Mais Stump suppose que cette possibilité alternative n'est pas contrôlée par l'agent : si elle se réalise, c'est à l'issue d'un phénomène physique indéterministe, un phénomène qui n'est pas lui-même voulu par l'agent. Puisqu'il est en amont de la volonté de l'agent, il n'est pas lui-même volontaire. Donc il n'est pas contrôlé par l'agent. Cela permet de conclure que les possibilités alternatives en question ne sont pas au contrôle de l'agent, et qu'elles ne sont donc pas pertinentes du point de vue de la responsabilité morale. Selon Stump, le libertarien ne peut pas les invoquer pour soutenir une version corrigée du principe des possibilités alternatives.

Widerker argumente d'une façon similaire. Les personnes-Z, placées devant la tentation du mensonge, ne peuvent pas dire la vérité. Soit elles mentent, soit elles s'évanouissent. Dans cet exemple, une personne-Z qui ment paraît bien responsable, car son acte de mentir n'était pas déterminé causalement. Or elle n'a pas de véritable possibilité alternative, précise Widerker : elle peut certes, s'évanouir, mais cette possibilité n'est pas en son pouvoir (elle ne lui est pas « accessible »). Bref, Widerker et Stump argumentent l'un et l'autre en présentant un exemple d'action découlant d'un processus indéterministe non contrôlé par l'agent. Celui-ci n'a pas la possibilité de choisir autrement que dans le monde actuel. La seule alternative qui

³⁹ McKenna, « Robustness, Control, and the Demand for Morally Significant Alternatives : Frankfurt Examples with Oodles, and Oodles of Alternatives », p. 204 in Widerker & McKenna (2006).

demeure physiquement ouverte ne lui est pas accessible, c'est pourquoi elle n'est pas « robuste ».

La première limite de cette argumentation est la suivante : le critère de « robustesse » utilisé par Stump et Widerker, à savoir la deuxième condition de McKenna, pourrait être considéré par le lecteur libertarien comme une condition de la responsabilité morale. Le libertarien peut réagir tout autrement que Stump et Widerker devant les cas qu'ils présentent : à partir du moment où il sait que le processus indéterministe situé à la source du choix n'est pas contrôlé par l'agent, il peut trouver évident que ce qui en résulte échappe à la responsabilité de l'agent. En somme, l'argumentation peut se heurter à un conflit d'intuitions. Nos intuitions touchant la responsabilité de l'agent dépendent de la manière dont l'exemple est présenté. Une présentation assez vague peut favoriser une intuition qui perd de sa force ou même se dissipe lorsque l'exemple est présenté de manière plus explicite ou plus détaillée. Je pense qu'un phénomène de ce genre peut produire dans les exemples de Stump et Widerker. En précisant que l'agent n'a pas de possibilité alternative proprement dite, Stump et Widerker nous suggèrent involontairement qu'il n'est pas responsable de ce qu'il fait, son action (ou son choix) étant le résultat inévitable d'un processus qu'il ne contrôle pas. En somme, les critères de pertinence des possibilités alternatives ne servent à rien dans la stratégie compatibiliste, s'ils sont en même temps des conditions de la responsabilité morale aux yeux du libertarien.

McKenna et Pereboom ont remédié à ce défaut en formulant une condition supplémentaire de la « robustesse » des possibilités alternatives (la condition n° 1 formulée plus haut), selon laquelle « l'alternative doit être moralement significative ». Il y a en effet des alternatives qui sont à la fois accessibles à l'agent et sans aucun lien avec la responsabilité. C'est ce que soutient Pereboom dans cet exemple :

[Tax Evasion] Joe hésite à réclamer une déduction de taxe pour la somme considérable qu'il a payée lors de l'inscription administrative de la maison qu'il vient d'acheter. Il sait qu'il est illégal de ne pas déclarer cela, qu'il ne sera probablement pas saisi par la justice, et que s'il l'est, il pourra plaider en disant qu'il ignorait cette loi. Admettons qu'il ait un désir très fort, mais non insurmontable, de favoriser son propre intérêt, même si cela a un coût pour les autres et a un caractère illégal. Admettons aussi qu'il soit un agent libre au sens libertarien. Au moment crucial, sa psychologie fait que dans cette situation, il lui faut des raisons d'ordre moral pour renoncer à tricher sur les taxes. [...] il est causalement nécessaire, pour qu'il renonce à tricher sur les taxes, qu'une raison morale lui apparaisse avec une certaine force. Une telle

raison morale peut lui apparaître soit involontairement, soit en découlant de son activité volontaire (par exemple lorsqu'il veut la considérer, ou lorsqu'il cherche à se la représenter clairement). Cependant, une raison morale lui apparaissant avec cette force n'est pas causalement suffisante pour qu'il évite de tricher sur les taxes.

Si cette raison morale lui apparaissait, Joe pourrait, avec son libre arbitre de type libertarien, choisir de faire cela ou se retenir de le faire (en l'absence de l'intervenant et de son appareil). Mais pour être sûr qu'il choisira de tricher sur ses taxes, un neurologue lui plante un appareil grâce auquel, s'il sent une raison morale apparaître à Joe avec la force requise, il stimulera son cerveau par un signal électrique de telle sorte qu'il choisirait de tricher. Dans la situation actuelle, aucune raison morale ne lui apparaît avec cette force, et il choisit de tricher sur ses taxes, tandis que l'appareil demeure inactif.⁴⁰

Dans cet exemple, l'agent a des possibilités alternatives, car il peut faire un effort d'attention (qu'il ne fait pas dans le monde actuel) pour savoir ce qu'il doit faire, mais cette possibilité alternative n'est pas moralement pertinente, car si elle se réalisait, ce ne serait pas suffisant pour que Joe fasse le bon choix. Selon Pereboom, Joe est donc responsable de son choix bien que les possibilités alternatives accessibles à lui ne soient pas moralement significatives.

Or ici, une fois de plus, la responsabilité de l'agent est loin d'être évidente. On peut y voir un exemple de ce que les théologiens appellent « l'ignorance invincible » : Joe connaît des raisons de ne pas tricher sur ses taxes, mais il ne peut pas connaître la vraie valeur de ces raisons, et c'est pour cela qu'il ne peut pas éviter de tricher. Or puisqu'il ne peut acquérir par lui-même une connaissance qui lui est nécessaire pour faire le bon choix, on ne peut pas le considérer comme coupable. De plus, le choix de Joe paraît déterminé causalement. En effet, l'absence d'une condition nécessaire au bon choix équivaut à la présence d'une condition suffisante pour que l'agent fasse le mauvais choix : si p est nécessaire pour que q , $\text{non-}p$ est suffisant pour que $\text{non-}q$. Cela est évident en vertu du *modus tollens* : $q \rightarrow p$, $\text{non-}p \rightarrow \text{non-}q$. Néanmoins, il est encore possible, ici, de distinguer « déterminer causalement » et « causer de manière déterministe » : l'ensemble des conditions décrites dans l'exemple déterminerait Joe à choisir de tricher, mais sans *causer* ce choix. Une troisième objection pourrait être que si Joe est responsable de son choix, il l'est par omission : pour n'avoir pas considéré suffisamment

⁴⁰ Derk Pereboom, *Living without Free Will*, Cambridge University Press, Cambridge, 2001, p. 19.

les raisons de ne pas tricher. Or on n'est pas responsable d'un événement par omission si on n'a rien pu faire d'efficace contre cet événement. Et tel est bien le cas ici : Joe avait la possibilité de considérer les raisons de ne pas tricher, mais cela aurait été insuffisant, et inefficace (puisque le résultat aurait été le même que dans le monde actuel). Joe n'est donc pas responsable de n'avoir pas pu empêcher la tricherie, même si c'est lui qu'il l'a faite.

Une quatrième objection est formulée par Pereboom lui-même : « on peut répondre que pour que Joe soit responsable de son action, sa psychologie morale doit avoir été établie par des choix cruciaux qu'il a faits lui-même et pour lesquels il avait des possibilités alternatives robustes »⁴¹. En somme, si Joe est responsable dans Tax Evasion, ce ne peut être que de manière *dérivée*. Le principe des possibilités alternatives resterait donc intact, puisque le choix de Joe serait le résultat d'inclinations qu'il aurait lui-même cultivées dans des situations où il avait la possibilité de développer d'autres inclinations. Mais selon Pereboom, cet argument ne tient pas si on précise l'exemple de la manière suivante :

[Tax Evasion – 2e partie] Joe a été élevé dans un contexte où tout le monde se comporte de manière égoïste, mais il a aussi développé son sens critique. Comme la plupart d'entre nous, il a d'abord accepté les valeurs de sa propre famille et s'y est montré fortement attaché. Ensuite, il a découvert d'autres valeurs mais ne les a pas adoptées. Pour qu'il abandonne les valeurs de sa famille, il était causalement nécessaire que lui apparaissent une autre valeur avec une certaine force, ce qui pouvait arriver involontairement, ou par l'effet de sa volonté. Si cela arrivait, il pourrait encore choisir de garder ces valeurs ou de les rejeter (en l'absence du neurologue et de son appareil). Mais le neurologue, connaissant toutes ces données de la psychologie de Joe, installe son appareil avant que Joe ne commence à exercer son sens critique sur ses propres valeurs morales, de telle sorte que si Joe accepte une valeur morale opposée à elles, l'appareil stimule électriquement son cerveau pour qu'il garde ses valeurs morales initiales. Mais dans le monde actuel, l'appareil n'intervient pas, car bien qu'il considère des raisons de changer de valeurs, ces raisons ne lui apparaissent jamais avec une force suffisante pour provoquer une réaction de l'appareil.

⁴¹ *Ibid.*, p. 22.

Selon Pereboom, Joe est moralement responsable, or sa psychologie ne s'est pas formée grâce à ses propres choix dans des situations où il avait des possibilités alternatives vraiment significatives. Il ne s'agit donc pas d'un cas de responsabilité dérivée. Ce dernier point est clairement établi, mais les autres objections sont toujours valables. D'abord, Joe n'est pas responsable car il est dans une situation d'ignorance invincible : quand une personne ne peut pas savoir qu'elle doit faire x , elle n'est pas responsable de ne pas faire x . Ensuite, on n'est pas responsable d'un événement E par omission si on ne pouvait rien faire d'efficace contre E . Joe n'est donc pas responsable de son ignorance, ni de ce qui en découle : son choix et son action, etc. Enfin, le simple fait que Joe n'ait jamais eu de possibilité alternative vraiment significative du point de vue moral exclut intuitivement sa responsabilité. Il n'a jamais eu la liberté de choisir quel genre de personne il serait. Ses inclinations ne dépendent pas de sa volonté, ou bien celle-ci est causalement déterminée par des faits qu'il ne contrôle pas : les propriétés de son cerveau, les tendances de son milieu social et familial, etc. On ne peut donc pas présenter un tel exemple en supposant que les libertariens y verront un cas de responsabilité morale. Pourquoi penser que Joe est responsable ? Probablement parce que son choix découle de ce qu'il est : de son tempérament, de son caractère, ou en tout cas de ses inclinations spontanées. Mais cela ne suffit pas pour tout libertarien, même si l'on précise, comme Pereboom dans cet exemple, que Joe a le libre arbitre au sens libertarien et exerce une activité volontaire indéterminée. Il en est ici comme dans l'exemple de Mele : le comportement de l'agent n'est pas entièrement déterminé causalement. Il a donc des possibilités alternatives, mais il n'en a pas relativement au choix dont Mele (ou Pereboom) le juge responsable. En l'occurrence, Joe n'a pas le choix entre tricher et ne pas tricher (ou garder ses valeurs initiales et ne pas les garder).

En résumé, les exemples compatibilistes se répartissent en deux catégories : une première où l'agent est effectivement responsable mais a eu des possibilités alternatives qui lui étaient accessibles et étaient significatives du point de vue moral, et une seconde catégorie dans laquelle le fait qu'il n'ait pas eu de telles possibilités alternatives jette un sérieux doute sur sa responsabilité. A cette deuxième catégorie appartiennent les exemples que nous avons vus dans les sections précédentes, c'est-à-dire ceux de Stump, Widerker, Mele et Robb, Hunt,

et enfin Pereboom⁴². Ces derniers exemples sont interprétés à partir d'un même présupposé, à savoir que si les inclinations de l'agent n'ont pas rencontré d'obstacle ou n'ont pas été détournées artificiellement avant le choix, l'agent est responsable de son choix. Plus précisément, il supposent que si le choix de l'agent est l'effet de ses inclinations spontanées, il en est responsable. J'ai déjà critiqué ce principe à propos de Stump et Widerker :

(AA) Lorsqu'un choix ou une action est l'issue d'un processus indéterministe, celui qui fait ce choix ou cette action le fait *par lui-même*.

Ce principe est contredit par un autre, plus intuitif, que je vais défendre dans le deuxième chapitre : le principe du transfert de non-responsabilité :

(TNR) Si S n'est pas responsable de x, et si S n'est pas responsable du fait que x cause y, alors S n'est pas responsable de y.

Dans « Tax Evasion », il est clair que Joe n'a pas la responsabilité *ultime* de ce qu'il fait. Cela suffit pour qu'il ne soit pas responsable aux yeux d'un libertarien. En effet, si un agent n'est pas responsable d'un événement ou d'un état de choses x, il n'est pas responsable non plus de ce qui en découle nécessairement. Il n'est donc pas responsable de ce qui en découle en vertu des lois de la nature. Or TNR s'applique aussi aux processus indéterministes : si je ne suis pas responsable de x, et si y est causé par x selon un processus indéterministe échappant entièrement à mon contrôle, je ne suis pas responsable de y. Dans « Tax Evasion », par exemple, le choix et l'action de Joe découlent de ses inclinations. Mais si ses inclinations viennent de son éducation, et indirectement, de faits appartenant à l'histoire de sa famille avant sa naissance, il ne peut être responsable de son choix. Cela vaut même si la relation causale conduisant de P à A est indéterministe. En effet, si le choix de Joe découle de ses inclinations par un processus causal indéterministe, mais qu'il ne contrôle ni ce processus ni ses inclinations, il n'est pas responsable de son choix. Je pense que TNR a une grande force intuitive. C'est pourquoi il faut le considérer comme valable *prima facie*, c'est-à-dire tant qu'il

⁴² Pereboom propose un exemple compatibiliste (« Tax Evasion ») mais sa position est l'incompatibilisme dur (cf. *ibid.*, p. 127 et s.), car le principe du transfert de non-responsabilité lui paraît solide. C'est pourquoi même si le principe des possibilités alternatives lui semble difficile à défendre, il ne peut adhérer au compatibilisme.

n'est pas réfuté par des contre-exemples. Or s'il est correct, il neutralise tous les contre-exemples au principe des possibilités alternatives qui appartiennent à la deuxième catégorie que je viens de mentionner, car tous reposent finalement sur le principe AA.

2.8 Résumé

Dans les sections précédentes, j'ai essayé de montrer que l'intuition incompatibiliste exprimée par le principe des possibilités alternatives est capable de résister face aux différents contre-exemples avancés par ses adversaires compatibilistes. En 2.2, nous avons vu que les exemples à la Frankfurt laissent à l'agent un « *flicker of freedom* ». Ce résidu des possibilités alternatives rend difficile l'interprétation des exemples à la Frankfurt. C'est pourquoi les compatibilistes ont essayé de trouver des exemples à la Frankfurt où l'agent n'a plus ni la possibilité d'agir autrement, ni celle de choisir autrement. En 2.3, j'ai discuté un exemple de ce genre chez Stump, or même dans ce cas, l'agent a la possibilité d'incliner vers différents choix avant qu'on intervienne au besoin pour déterminer son choix.

Pour éviter que cela soit considéré comme un *flicker of freedom*, Stump a recours à des hypothèses annexes qui compliquent encore l'interprétation de l'exemple en la faisant dépendre d'une certaine théorie de l'identité psycho-physique. Widerker résout le problème en imaginant un exemple sans intervenant contrefactuel (et sans « signe ») mais cet exemple a deux défauts. Premièrement, comme Stump, Widerker précise que les possibilités alternatives de son exemple ne sont pas accessibles à l'agent. À cause de cette précision, un libertarien peut estimer que l'agent n'est pas responsable de ce qu'il fait. Ainsi, l'exemple n'atteint pas son but, puisqu'il ne s'agit plus d'un cas de responsabilité morale. Deuxièmement, on peut se demander si le scénario imaginé par Widerker est cohérent. Dans « Z-personnes », l'action de Jones est issue d'un processus indéterministe, mais l'autre possibilité pouvant se réaliser par ce même processus ne lui est pas accessible. Le problème est alors le suivant : puisque Jones est responsable dans le monde actuel, il a accès à la possibilité qui se réalise dans le monde actuel. Jones a donc accès à la possibilité qui se réalise dans le monde actuel. Or l'autre possibilité, si elle se réalisait, serait l'issue du même processus indéterministe. Jones devrait donc y avoir également accès, contrairement à ce que suppose l'exemple.

Nous avons vu ensuite que pour éviter ces objections, il faut présenter un cas où il y a des possibilités alternatives accessibles à l'agent, mais bloquées au moment précis du choix. Cette stratégie adoptée par Mele et Robb pose également problème. Pour un libertarien, le fait que l'agent ait des possibilités alternatives avant le moment du blocage importe peu puisqu'elles n'existent plus au moment pertinent, et cette situation suggère que le choix de l'agent est causalement déterminé. Cette idée est d'ailleurs confirmée par le fait que le processus de blocage agit dans le monde actuel et pas seulement dans des mondes possibles.

Face à l'échec des stratégies de blocage, la dernière solution possible est d'imaginer des exemples où les possibilités alternatives ne sont pas supprimées au moment crucial, et où elles sont accessibles à l'agent (pour éviter l'inconvénient des exemples de Stump et Widerker) tout en montrant qu'elles ne sont pas moralement pertinentes. Cette stratégie adoptée par McKenna et Pereboom est peu convaincante elle aussi, car les critères de « robustesse » ou de pertinence que l'on peut avancer pour montrer que les possibilités alternatives en question sont négligeables, peuvent être aux yeux du libertarien des conditions de la responsabilité morale elle-même. Ainsi, comme dans les deux autres types d'exemples, on peut aboutir à un conflit d'intuitions où les libertariens estiment que l'agent n'est pas responsable.

Frankfurt a montré que pour être acceptés par des libertariens, les contre-exemples doivent être des situations où un individu S est responsable de ce qu'il fait, où son action est pourtant rendue inévitable par certains facteurs, et où ceux-ci ne *causent pas de manière déterministe* l'action de S. Si cette dernière condition n'est pas remplie, le fait que les facteurs causent l'action de manière déterministe suffit pour qu'un libertarien confronté à ces exemples n'accepte pas de considérer l'agent comme moralement responsable, et ainsi l'exemple ne parvient pas à réfuter le principe des possibilités alternatives. Or même s'il est effectivement possible de présenter des cas de ce genre (comme l'ont soutenu Mele et Robb, Hunt, Stump, et Widerker lui-même), ces exemples ne paraissent pas assez convaincants. Nous avons vu pourquoi en examinant les différentes réponses à l'argument du dilemme. En définitive, je pense que ces exemples sont trop complexes pour mettre en évidence nos intuitions sur la responsabilité morale. Ils ne sont pas seulement irréels, mais d'une complexité telle que nous ne voyons pas assez bien, en les lisant, si l'agent est vraiment responsable de ce qu'il fait. Dans certains cas, la cohérence de l'exemple est douteuse ; et dans tous ceux où on cherche à

éliminer le *flicker of freedom* sans supposer le déterminisme, la responsabilité de l'agent est douteuse. Finalement, le libertarien peut avoir le sentiment que l'agent n'est responsable dans aucun de ces cas. La responsabilité de l'agent suppose qu'il ait accès à la possibilité d'agir autrement ou qu'il ait eu accès auparavant à une possibilité alternative lui permettant d'agir autrement au moment en question. Cette condition doit être ajoutée à celles qui définissent les situations IRR. En effet, si elle n'est pas remplie, l'exemple tombe sous le coup de l'argument du transfert ou « argument de la conséquence ».

Dans les prochaines sections, j'examinerai les différentes versions de l'argument du transfert⁴³. Je soutiendrai que cet argument est correct et que parmi ses différentes formulations, celles qui ont été réfutées par des contre-exemples sont toujours plus ou moins éloignées de l'idée que veut exprimer l'argument. Inversement, les formulations les plus fidèles à l'idée de départ semblent irréfutables.

⁴³ Van Inwagen défend cet argument dans *An Essay on Free Will*, p. 182-188.

III L'argument du transfert

3.1 Introduction. L'argument du transfert (ou de la « conséquence »)

Le déterminisme est la thèse selon laquelle tout événement est une conséquence nécessaire d'événements passés en vertu des lois de la nature. Il en découle la nécessité de tout événement, car les événements passés sont nécessaires à partir du moment où ils appartiennent au passé. Notons « L » l'ensemble des lois de la nature, « P » l'ensemble des événements passés, et « F » l'ensemble des événements présents et futurs. L'expression « Np » signifiera : « p et personne ne peut empêcher que p »¹. Dans cette expression, le verbe « peut » a un sens intemporel : il signifie que personne ne peut faire, n'a pu faire, ni ne pourra faire quelque chose qui empêche p de se produire. Nous pouvons penser que si le déterminisme est vrai, le raisonnement suivant est correct :

1. NP
2. NL
3. N(P et L) (en vertu de 1 et 2). Or :
4. N((P et L) → F) Donc :
5. NF (en vertu de 3 et 4)

Ce qui découle nécessairement de propositions nécessaires est à son tour nécessaire ; c'est une règle en logique modale :

Si $\Box a$
 et si $\Box (a \rightarrow b)$
 alors $\Box b$

Puisque le connecteur N exprime une sorte de nécessité, on peut supposer que cette règle est aussi valable pour lui, et justifie le passage des propositions (3) et (4) à la conclusion.

¹ Je reprends la définition de Van Inwagen dans *An Essay on Free Will*, p. 93.

Ainsi, comme nul n'est responsable de ce qui est inévitable, le déterminisme exclut la responsabilité morale. Si nous sommes responsables d'une partie de nos actions, il est faux que $(P \text{ et } L) \rightarrow F$. On voit ainsi comment l'argument du transfert soutient le principe des possibilités alternatives : si la responsabilité morale des agents humains exclut l'idée que $(P \text{ et } L)$ rend nécessaire F , elle implique que dans certaines situations, nous avons la possibilité d'agir autrement.

Dans les prochaines sections, j'examinerai les différentes stratégies compatibilistes destinées à fragiliser l'argument du transfert. Elles sont au nombre de quatre. Il est d'abord possible de rejeter une des prémisses : en niant la fixité du passé (1), ou celle des lois (2). La troisième prémisses est (4). Or elle exprime la thèse déterministe. On ne peut donc critiquer l'argument en la rejetant. En revanche, il est possible de contester le passage de (1) et (2) à (3), ou le passage de (3) et (4) à (5). Ces inférences sont valides avec l'opérateur « $\square$ » à la place de « N ». Mais on peut se demander si l'opérateur « N » est agglomératif et clos sous l'implication matérielle. Si ces deux conditions ne sont pas réunies, l'inférence ne peut être valide. Une quatrième façon de rejeter l'argument est de dire que même si l'inférence sur laquelle il se fonde est probablement valide, elle ne réfute pas la thèse compatibiliste, car comme le mot « capable » a plusieurs sens, l'inévitabilité de F n'empêche pas de penser qu'*en un certain sens*, l'agent a la capacité de faire que non- F .

J'examinerai ces quatre façons de réfuter l'argument du transfert ou d'en limiter la portée. En 3.2 et 3.3, je défendrai les principes de fixité du passé (FP) et des lois de la nature (FL). En 3.4, j'aborderai la thèse de Michael Slote qui met en doute la validité de l'argument en soutenant que l'opérateur « N » n'est pas agglomératif. En 3.5, je discuterai la dernière stratégie.

Toutes ces stratégies compatibilistes s'appuient sur des contre-exemples pour réfuter le principe du transfert d'incapacité (TI). Je soutiendrai qu'elles échouent. Mais un compatibiliste peut encore rejeter le principe ce principe en déclarant qu'il n'est pas évident et en montrant qu'il n'y a aucun moyen de le prouver. J'examinerai donc, en 3.6, s'il est possible de le prouver en le dérivant d'un autre principe. Cela m'amènera à défendre plusieurs versions du transfert d'incapacité (TI). Je pense que même si elles étaient incorrectes, il faudrait accepter TNR, le

transfert de non-responsabilité, principe le plus immédiatement pertinent dans le débat sur le déterminisme et la responsabilité morale. Pour finir, en 3.7 et 3.8, je défendrai donc TNR contre les critiques qui le visent directement.

3.2 La fixité des lois de la nature (FL)

Dans l'argument du transfert, les prémisses (1) et (2) reposent sur l'idée que le passé et les lois de la nature sont fixes et échappent entièrement au contrôle des hommes. L'argument repose donc sur le principe de la fixité des lois de la nature (FL) et le principe de la fixité du passé (FP). Définissons ces principes et essayons d'en estimer la solidité. Le premier fait appel à la notion de lois de la nature, très difficile à définir, mais quelle que soit la bonne manière de comprendre les lois de la nature d'un point de vue ontologique, l'idée exprimée par FL a une signification assez claire. Pour Van Inwagen, « une loi est une proposition »² et le fait pour une proposition d'être une loi de la nature n'est pas une caractéristique qui dépend de l'état actuel de la science ou de l'histoire des connaissances scientifiques. On peut aussi concevoir les lois de la nature comme des caractéristiques permanentes de la nature qui sont toujours traduites dans le langage sous une forme propositionnelle. Par commodité, je m'en tiendrai à la première conception, sachant que si elle est fautive, nous pouvons peut-être corriger la formulation du principe sans qu'il perde une partie de sa force intuitive.

Van Inwagen distingue deux versions de FL : le principe *de dicto* et le principe *de re*. Selon le principe *de dicto*, « il est impossible qu'il y ait une personne x et une proposition y telles que x puisse faire que « y est une loi de la nature » et « y est fautive » ». Cette affirmation paraît difficilement contestable, mais le principe dont nous avons besoin est le principe *de re* :

(FL) Il est nécessaire, pour toute personne x et toute proposition y , que si y est une loi de la nature, alors x ne peut rendre fautive y .

² Peter Van Inwagen, *An Essay on Free Will*, p. 60.

Par exemple, si c'est une loi de la nature que « tout aimant attire le fer », Pierre-André Rousseau ne peut rendre fausse cette proposition. Cela découle du principe *de re* mais non du principe *de dicto*. Il faut noter qu'un tel raisonnement est correct même si les lois de la nature sont des propositions contingentes. Il y a des mondes possibles où Pierre-André existe, où cette proposition n'est pas une loi de la nature et où Pierre-André est capable de la rendre fausse. Mais cela n'empêche pas que dans le monde actuel et dans tous les mondes possibles où cette proposition est une loi de la nature, Pierre-André est incapable de la rendre fausse.

Si l'on donne un sens très large à l'expression « loi de la nature », il est sans doute possible de violer une loi de la nature. La psychologie, la sociologie, la linguistique et l'anthropologie énoncent des lois du comportement humain que nous avons la liberté de ne pas suivre. Par exemple, Van Inwagen imagine³ que les psychologues découvrent une loi selon laquelle tout individu ayant reçu une éducation du type A pendant son enfance répand des rumeurs calomnieuses sur ses collègues de travail. Si cette loi existait, elle ne voudrait pas dire que les personnes ayant reçu une éducation du type A seraient incapables de parler plus honnêtement de leur collègues. Elle décrirait une tendance plutôt qu'une régularité stricte dans le comportement d'une catégorie de personnes. Pour la suite, nous laisserons de côté ce genre de lois et nous prendrons le principe FL dans son sens restreint, en réservant le nom de « lois de la nature » aux propositions qui s'appliquent à des choses dépourvues de raison telles que les électrons, les planètes et les choses fabriquées par l'homme. Cela doit suffire car la thèse déterministe est généralement comprise comme l'idée que tout événement est causalement déterminé par des processus physiques ou biologiques. Ces définitions une fois posées, il semble logiquement nécessaire que les lois de la nature sont inviolables. On peut concevoir, dit Van Inwagen, que les pouvoirs de l'homme augmentent au point qu'il deviendra un jour capable de dévier la trajectoire de certains astres. Mais imaginons la situation suivante :

[...] un bureaucrate du futur ordonne à un ingénieur de construire un vaisseau spatial capable de voyager plus vite que la lumière. L'ingénieur lui dit qu'il y a une loi de la nature selon laquelle rien ne peut aller plus vite que la lumière. Le bureaucrate admet cette difficulté, mais lui conseille de persévérer : « je suis

³ Peter Van Inwagen, *An Essay on Free Will*, p. 63-64.

sûr que si vous travaillez beaucoup et avec une grande intelligence, vous trouverez un moyen d'aller plus vite que la lumière, même si c'est une loi de la nature que rien ne peut le faire.⁴

Comme le dit Van Inwagen, cette demande est clairement incohérente. Cependant, David Lewis a montré qu'en un certain sens, nous pouvons agir de telle sorte que les lois de la nature sont contrées :

Le déterminisme doux semble avoir une conséquence incroyable. Il semble impliquer, en vertu de certaines prémisses supplémentaires acceptables, que nous sommes parfois capables d'agir de telle sorte que les lois de la nature sont contrées. Mais si nous distinguons une version forte et une version faible de cette conséquence incroyable, nous trouverons à mon avis que c'est la version forte qui est incroyable et que c'est la version faible qui est la conséquence.⁵

Le déterminisme doux est la thèse selon laquelle bien que nos actions soient des conséquences nécessaires d'événements passés en vertu des lois de la nature, nous avons souvent la capacité d'agir autrement. Par exemple, je pose ma main sur mon bureau, mais j'aurais pu lever mon bras. Pourtant, si le déterminisme est vrai, mon geste est une conséquence nécessaire du passé actuel et des lois de la nature. Avais-je donc la possibilité de lever mon bras ? Appelons « P » l'ensemble des propositions énonçant des faits passés et « L » l'ensemble des propositions énonçant les lois de la nature. L'affirmation que je pouvais lever mon bras au lieu de le poser sur le bureau suppose qu'une des trois propositions suivantes soit vraie :

1. Les propositions contraires peuvent être vraies en même temps.
2. Il y a une proposition dans P qui est fausse (il y a donc un fait passé qui ne s'est pas produit).
3. Il y a une proposition dans L qui est fausse.

Quelle proposition faut-il admettre ? Sûrement pas la première, puisqu'elle est une négation du principe de non-contradiction. La deuxième est également fausse car le passé ne peut être changé. Il reste donc la troisième, mais encore faut-il lui donner un sens acceptable.

⁴ Peter Van Inwagen, *An Essay on Free Will*, p. 62.

⁵ David Lewis, "Are we free to break the laws ?", *Theoria* 47 (1981), p. 113-121. Ce passage se trouve p. 122 dans l'anthologie de Gary Watson (2003).

Bien entendu, précise Lewis, tant qu'une loi est une loi, elle ne peut être violée. L'idée de Lewis est donc qu'un énoncé général peut cesser momentanément d'être une loi. Voici maintenant les deux versions de la thèse :

(Version faible) : Je suis capable d'agir de telle sorte que si j'agissais ainsi, une loi aurait été violée.

(Version forte) : Je suis capable de violer une loi.

Comme le dit Van Inwagen⁶, personne n'est capable de violer les lois de la nature. Par exemple, nous ne sommes pas capable d'aller plus vite que la lumière. Mais en donnant cet exemple, Van Inwagen avait à l'esprit la version la plus forte de l'expression « rendre fausse une loi de la nature ». Lewis envisage un autre cas : nous pouvons imaginer un miracle permettant que je lève mon bras au lieu de le poser sur mon bureau. Dans cet exemple, je ne suis pas capable de violer une loi de la nature, et les lois de la nature sont les mêmes que dans le monde actuel. Il y a seulement eu un petit miracle avant mon action.

Cet argument peut sembler *ad hoc* et excessivement étrange. Pour le comprendre, il faut y voir une application de la théorie de Lewis sur les mondes possibles et les conditionnels contrefactuels⁷. Le monde possible où il y a un miracle ponctuel, c'est-à-dire un événement qui viole les lois de la nature du monde actuel, est beaucoup plus proche du monde actuel que celui où je possède des capacités complètement différentes (ou bien celui où les lois de la nature ne sont pas les mêmes). Dire qu'il y a eu un miracle à un moment précis, ce n'est pas dire que je sois au-dessus des lois de la nature. Par contre, si nous pouvions aller plus vite que la lumière, nous serions affranchis d'une loi physique, idée manifestement absurde. La version faible n'a donc pas l'inconvénient de la version forte. Dans celle-ci, je suis capable de faire un miracle alors que dans la version faible, je suis seulement capable de faire une action qui suppose un miracle à un moment antérieur. Cela a donc du sens, si je pose la main sur mon bureau à t_2 , de dire que je pourrais lever mon bras à t_2 même si le déterminisme est vrai.

⁶ Peter Van Inwagen, *An Essay on Free Will*, p. 61-62.

⁷ Cf. David Lewis, *Counterfactuals*, Basil Blackwell, Oxford, 1973.

Il faut donc distinguer deux interprétations du principe de fixité des lois (FL), une forte et une faible. La deuxième est l'interprétation compatibiliste du principe de fixité des lois :

(FLnc) Pour tout agent S, et action A, si l'action A implique la fausseté de L (conjonction de toutes les lois de la nature du monde actuel), alors S ne peut faire A.

(FLc) Pour tout agent S, événement E, action A, et pour tous instants t_1 , t_2 et t_3 ($t_1 \leq t_2 \leq t_3$), si (1) A, à t_2 , est compatible avec L, et si (2) A, à t_2 , causerait à t_3 un événement E incompatible avec L, alors S ne peut pas, à t_1 , accomplir l'action A.⁸

Selon Lewis, nous n'avons pas la capacité de faire une action incompatible avec L ou qui causerait un événement incompatible avec L, mais rien n'empêche de penser que nous pouvons faire des actions qui impliquent à un moment antérieur une violation des lois du monde actuel. En effet, cela n'est pas contraire à **FLc**.

On peut se demander néanmoins si le verbe pouvoir exprime ici une capacité de l'agent ou bien seulement une possibilité dont la réalisation ne dépend pas de lui. Nous avons vu dans le premier chapitre que le verbe « pouvoir », même appliqué à une personne, ne désigne pas toujours une capacité. C'est le cas dans des phrases comme : « je pourrais tomber de cet arbre et me blesser », ou « je pourrais faire une allergie à ce médicament ». Le fait qu'un miracle se soit produit avant mon action ne dépend pas de mon pouvoir et de ma volonté. On ne peut donc pas dire que j'ai eu le pouvoir (ou la capacité) de lever mon bras à t_2 si cette action suppose un miracle au préalable.

Ce que veut dire Lewis est qu'il y a un monde possible proche du monde actuel et dans lequel je lève mon bras à t_2 au lieu de le poser sur mon bureau. Mais cette possibilité ne paraît pas suffisante pour que j'aie le pouvoir ou la capacité de lever mon bras à t_2 . Encore faut-il que ce monde me soit accessible.

⁸ Je m'inspire ici des formules utilisées par Fischer à propos de la fixité du passé, mais en les adaptant pour exprimer une distinction analogue entre deux versions du principe de fixité des lois. Cf. J. M. Fischer, *The Metaphysics of Free Will*, p.78-79.

Inversement, on peut montrer que si nous avons la capacité de violer les lois de la nature au sens faible, comme le soutient Lewis, nous devrions aussi avoir celle de violer les lois de la nature au sens fort, ce que nie Lewis. En d'autres termes, le compatibilisme du miracle local n'a aucun moyen de justifier sa distinction entre les capacités faible et forte de violer les lois de la nature. Cette objection que je vais exposer maintenant a été élaborée par Helen Beebe⁹.

Reprenons la distinction de Lewis. Elle suppose qu'il y a deux manières pour une action non actuelle de violer les lois de la nature : pour tout agent S et toute action A, A viole les lois de la nature si et seulement si :

- Si S fait A, une loi est violée, mais ni par A ni par un effet de A (S a alors la capacité *faible* de violer les lois de la nature).
- Si S fait A, une loi est violée par A ou par un effet de A (S a alors la capacité *forte* de violer les lois de la nature).

Lewis définit ainsi un événement violant une loi (« *law-breaking event* »):

(LBE1) Un événement E viole une loi (du monde w) si et seulement si, nécessairement, si E se produit, alors la proposition L (conjonction des lois de la nature de w) est fausse.¹⁰

Cette définition n'est pas satisfaisante, car de nombreux événements violant des lois ne correspondent pas à cette définition. Par exemple, supposons que dans le monde actuel, je reste assis à mon bureau pendant la prochaine demie heure. Quel est, entre les deux mondes possibles suivants, le plus proche du monde actuel ? Dans w1, je reste assis à mon bureau pendant 9 mn 59 s, le cours des événements étant exactement le même que dans le monde actuel, puis je disparaîs soudain pour réapparaître 1 seconde plus tard à l'entrée d'un bar. Dans w2, je reste à mon bureau pendant 5 mn, le cours des événements étant le même que dans le monde actuel, puis j'ai soudain le désir de boire une bière. Un peu plus tard (disons 5 mn après), je cède à ce désir et me rends au bar.

⁹ Helen Beebe, "Local Miracle Compatibilism", *Noûs* 37/2 (2003), p. 258-277.

¹⁰ Cf. *ibid.*, p. 265.

Si le déterminisme est vrai, w_1 et w_2 contiennent tous deux des miracles relativement au monde actuel. Mais l'événement E (« j'entre dans le bar au bout de 10 mn ») semble moins miraculeux dans w_2 que dans w_1 , qui est manifestement miraculeux. Or d'après la définition (LBE1), E-dans- w_1 n'est pas un miracle. En effet, la définition stipule qu'un événement viole les lois de w seulement s'il les viole dans tous les mondes possibles. Ce n'est pas le cas de E-dans- w_1 , car si cet événement viole L dans tous les mondes possibles où les circonstances de E sont les mêmes que dans le monde actuel, on ne peut pas dire qu'il viole L dans tous les autres mondes possibles quelles que soient les circonstances. Le fait qu'un événement viole L ne dépend pas seulement de la nature de l'événement lui-même, mais aussi des circonstances dans lesquelles il se produit. Il faut donc reformuler (LBE1) de la manière suivante :

(LBE2) Un événement E dans le monde w_a viole les lois du monde w_b si et seulement si E, dans les circonstances où il se produit dans w_a , est incompatible avec L (défini comme la conjonction de toutes les lois de w_b).¹¹

Comme nous l'avons vu, le compatibilisme du miracle local est la thèse selon laquelle nous avons dans certains cas la capacité faible, et non la capacité forte, de faire une action A qui suppose un événement-violant-les-lois (à un moment antérieur à A), mais nous n'avons pas la capacité de faire une action qui viole elle-même les lois ou dont un effet viole les lois. En somme, le miracle (c'est-à-dire l'événement violant les lois du monde actuel) ne doit pas être l'action elle-même ou un de ses effets. Si je n'ai pas levé le bras dans le monde actuel, « le miracle de divergence (*divergence miracle*) ne pourrait pas être mon acte de lever la main. Cet acte est complètement absent de la série des événements actuels, donc il ne peut pas apparaître à moins qu'il y ait déjà eu une divergence... C'est pourquoi il ne peut pas être un de mes actes ».¹²

Cette affirmation paraît difficile à justifier. En effet, s'il est possible qu'il y ait, avant t , un événement contraire aux lois de la nature et permettant que je lève la main à t , pourquoi serait-il impossible qu'il y ait à t un événement également contraire aux lois de la nature et

¹¹ Cf. *ibid.*, p. 266.

¹² David Lewis, "Are we free to break the laws ?", p. 294 dans l'anthologie de Gary Watson (2003).

consistant dans le fait que je lève la main ? Pour Lewis, le miracle de divergence doit se situer avant l'action non-actuelle (par exemple lever la main) car il ne peut être l'action elle-même, et il ne pourrait pas entraîner l'action s'il lui était simultané.

Mais en admettant que le miracle de divergence soit en effet antérieur à M, il faut lui appliquer le même raisonnement : cet événement ne peut pas violer *au sens fort* les lois de la nature ; il les viole donc au sens *faible*. Il n'est donc pas lui-même le miracle de divergence, mais implique un miracle de divergence situé avant lui, et ainsi de suite. Ce raisonnement entraîne dans une régression à l'infini.¹³ Si le miracle de divergence se situe avant l'action M, il ne faut pas qu'il consiste en une autre action simplement antérieure à M, car cela impliquerait que l'agent ait exercé (avant M) une capacité *forte* de violer les lois de la nature. Il faut donc que le miracle ne soit pas une action mais un simple événement.

Finalement, selon cette idée, la possibilité d'un événement contraire aux lois n'est pas évaluée en fonction de sa proximité ou de son éloignement par rapport au monde actuel, mais au moyen de la distinction entre les actions et les autres événements. En effet, considérons l'événement non-actuel M, ma décision de lever ma main. Il n'y a aucune raison de supposer que le monde possible le plus proche du monde actuel et dans lequel je lève la main ne puisse pas être un monde où cette action elle-même soit le miracle de divergence.

Le compatibilisme du miracle local suppose ainsi que les contraintes imposées par les lois de la nature soient plus fortes pour les actions que pour les événements qui ne sont pas des actions¹⁴. Or cette idée est étrangère aux conceptions les plus communes des lois de la nature. Nous pouvons distinguer deux conceptions principales des lois de la nature. Selon la première, que l'on peut qualifier de « nécessitariste », les lois de la nature ne sont pas inviolables simplement au sens où il est conceptuellement impossible qu'elles soient violées, mais parce qu'elles ont une sorte de réalité et qu'elles agissent sur les choses. Selon une autre conception, que l'on peut qualifier de « humienne », les lois de la nature sont seulement des faits ou des propositions au sujet des régularités (selon une version améliorée de cette idée, la

¹³ Cf. Helen Beebe, "Local Miracle Compatibilism", p. 270.

¹⁴ Cf. Helen Beebe, "Local Miracle Compatibilism", p. 273 et s.

conception de Ramsey-Lewis, elles sont plutôt un sous-ensemble de régularités : ce sont les régularités qui expliquent le plus de faits). Ainsi comprises, les lois de la nature n'imposent aucune contrainte aux choses et aux événements ; au contraire : c'est le comportement des choses et la succession des événements qui détermine les lois. Mais malgré toutes leurs différences, ces deux conceptions des lois de la nature s'accordent sur ce point crucial : dans chacune de ces conceptions, les lois contraignent autant (et dans le même sens du mot) les actions et les événements. Dans aucune de ces conceptions, elles n'imposent aux actions une limite qu'elles n'imposent pas aux autres événements. De son côté, le compatibilisme du miracle local implique une conception hybride des lois de la nature. Il affirme que les agents ont parfois la capacité (faible) d'agir de telle sorte que les lois sont violées, affirmation qui, pour être plausible, suppose une conception humienne des lois de la nature, où elles ne sont que des propositions exprimant des régularités. Mais il déclare en même temps qu'aucune action ne peut violer les lois de la nature, ce qui suppose plutôt une conception nécessitariste des lois de la nature.

La thèse de Lewis paraît donc indécise en ce qui concerne la nature des lois. Pour cette raison, le compatibilisme du miracle local ne semble pas viable et ne parvient pas à mettre en doute l'argument du transfert.

Je pense qu'il a encore un défaut. En effet, le miracle de divergence n'est pas seulement un événement violant les lois de la nature, mais aussi un événement qui ne s'est pas produit dans le monde actuel. Aussi, en affirmant comme Lewis que nous pouvons faire à t une action impliquant un « miracle local » avant t , on suppose que le passé pouvait être différent, car le miracle en question ne fait pas partie du passé actuel à t . La capacité de S de faire l'action A à t suppose donc (à t) que le passé pouvait être différent. Ainsi, la thèse de Lewis suppose de redéfinir non seulement FL, mais aussi FP (la fixité du passé). Il paraît moins coûteux de réviser seulement un des deux principes de fixité. Mais pour contester FL d'une autre manière que Lewis, c'est-à-dire sans avoir besoin de contester FP en même temps, il faut soutenir que nous pouvons faire des actions telles que certaines lois du monde actuel seraient globalement fausses, et non seulement suspendues pendant un bref moment antérieur à l'action. Cela suppose que nous ayons accès à des mondes possibles très différents du monde actuel. Nous voyons donc que la remise en cause de la fixité des lois de la nature a de toute façon des

conséquences peu plausibles. Nous allons voir maintenant que l'autre principe de fixité, le principe de fixité du passé, a lui aussi une grande solidité, et qu'il est difficile de le reformuler de manière à rendre plausible la thèse compatibiliste.

3.3 La fixité du passé (FP)

Considérons maintenant le principe de fixité du passé utilisé dans l'argument de la conséquence :

(FP) Pour toute action Y, tout événement E, tout agent S, et tout instant t, s'il est vrai que (si S faisait Y à t, un fait du passé relatif à t n'aurait pas été un fait), alors S ne peut pas faire Y à t.¹⁵

Comme FL, le principe de fixité du passé peut s'interpréter de deux manières, une causale (FPc) et une non-causale (FPnc) :

(FPc) Pour toute action y, tout événement E, tout agent S et tout instant t, s'il est vrai que (si S faisait Y à t, S initierait une séquence causale entraînant la non occurrence d'un événement E qui s'est produit actuellement dans le passé relatif à t), alors S ne peut pas faire Y.

(FPnc) : Pour toute action Y, tout événement E, tout agent S et tout instant t, s'il est vrai que (si S faisait Y à t, un événement qui s'est produit actuellement dans le passé relatif à t ne se serait pas produit), alors S ne peut pas faire Y.

FPc, comme FLc, est difficilement contestable. En revanche, FPnc et FLnc, les deux versions non-causales des principes de fixité, font l'objet d'un débat entre les compatibilistes et les incompatibilistes. Nous avons vu Lewis critiquer FLnc en montrant qu'il y a des mondes possibles où l'agent peut faire Y bien que cela suppose une violation des lois de la nature. De

¹⁵ Je reprends ici les formules de Fischer dans *The Metaphysics of Free Will*, p. 78-79.

la même façon, Fischer et Saunders exposent des cas où un agent S peut faire Y bien que cela suppose un passé différent. Ces exemples ont la structure suivante :

- (a) A t2, S ne fait pas Y.
- (b) Si S faisait Y à t2, alors E se serait produit à t1.
- (c) E ne s'est pas produit à t1.
- (d) Pourtant, S peut faire Y à t2.

La deuxième proposition est un conditionnel allant d'un événement à un autre événement antérieur au premier. Dans les circonstances décrites par (a), et (c), ce conditionnel paraît incompatible avec (d), c'est-à-dire l'attribution de capacité. Or voici un exemple de Fischer où ces deux propositions semblent vraies en même temps :

[Vieux loup de mer] « Chaque matin à neuf heures (depuis quarante ans), le vieux loup de mer appelle le service météo pour s'informer du temps qu'il va faire à midi. Si on lui répond qu'il va faire beau à midi, il sort en mer à midi. Mais jamais il ne sort en mer à midi si on lui a dit qu'il ferait mauvais à midi. Le vieux loup de mer a un comportement très régulier et des dispositions psychologiques parfaitement stables – il fait attention aux prévisions météorologiques, il n'est pas négligent, confus ou psychologiquement instable, et alors qu'il aime beaucoup naviguer sous le soleil, il déteste naviguer par gros temps. »¹⁶ On suppose en outre que le comportement du vieux loup de mer n'est pas déterminé causalement, mais explicable et prévisible à partir de ses valeurs, de ses désirs et de ses croyances.

« Il est maintenant midi. A 9 heures, le vieux loup de mer a téléphoné au service météo qui lui a annoncé un très mauvais temps pour midi : beaucoup de vent et des pluies torrentielles. Le vieux marin est en pleine forme, son bateau est prêt pour la navigation. Mais parce qu'il a en tête les prévisions météo, il décide à midi de ne pas sortir en mer. Or est-il incapable de naviguer cet après-midi ? Étant donné qu'il n'est ni contraint, ni hypnotisé, ni manipulé électroniquement, ni déprimé, etc. (et que le déterminisme

¹⁶ *Ibid.*, p. 80-81.

causal est faux), il est certainement *capable* de sortir en mer à midi. Simplement, il ne le fait pas : il choisit tout à fait rationnellement de ne pas le faire bien qu'il en ait le *pouvoir*. »¹⁷

En résumé,

- (a) A midi, le vieux loup de mer ne sort pas en mer.
- (b) Si le vieux loup de mer sortait en mer à midi, alors on ne lui aurait pas annoncé à 9 heures du mauvais temps pour midi.
- (c) A 9 heures, on lui a annoncé du mauvais temps pour midi.
- (d) Pourtant, le vieux loup de mer *peut* sortir en mer à midi.

Il semble que le conditionnel (b) et l'attribution de capacité (c) soient vrais en même temps. Nous aurions donc ici un contre-exemple à FPnc. Mais l'interprétation de cet exemple est problématique. Un libertarien pourrait douter de sa cohérence, car Fischer suppose que la décision du vieux loup de mer est déterminée par son caractère, or on peut se demander si la décision du vieux loup de mer peut être déterminée sans être déterminée causalement. Certes, l'attribution de capacité semble vraie malgré la vérité du conditionnel (b), et même si Fischer n'excluait pas le déterminisme causal dans cet exemple, le vieux loup de mer semblerait encore capable de sortir en mer à midi, car Fischer a précisé que le vieux marin n'est pas soumis à une contrainte, à un trouble psychique ou à une quelconque manipulation. Cette précision nous pousse à dire qu'il a effectivement le pouvoir de sortir en mer.

Mais si un libertarien lit l'exemple en y introduisant un déterminisme causal aboutissant à la décision du marin, il pourrait corriger son jugement en renonçant à l'attribution de capacité, car le déterminisme causal implique que quelque chose empêche le vieux loup de mer de prendre une autre décision. Une cause externe ou interne à l'agent, ou un défaut externe ou interne l'empêche de décider à midi d'aller en mer. Cela implique qu'il n'en est pas capable.

¹⁷ *Ibid.*, p. 81.

D'où l'importance de la précision faite par Fischer : il faut que dans l'exemple, le déterminisme causal soit faux (ou au moins, que la décision du marin ne soit pas déterminée causalement)¹⁸. Or en admettant que cette précision ne rende pas l'exemple incohérent (un point que je ne discuterai pas ici), elle ne suffit pas à convaincre le lecteur libertarien du fait que le vieux marin est bien capable de sortir en mer à midi. En effet, si le libertarien croit à la possibilité d'un déterminisme non-causal mais rationnel aboutissant à la décision de ne pas naviguer, il peut encore refuser l'attribution de capacité si, à ses yeux, le déterminisme non-causal imaginé dans l'exemple exclut que le marin ait la capacité de prendre une autre décision. Ainsi, quelle que soit la nature de ce déterminisme, il fait que le vieux marin est incapable de sortir en mer à midi. L'interprétation de l'exemple dépend de ce que l'on considère comme des conditions de la capacité d'agir autrement. Si un agent S fait Y à t2, on peut penser qu'il a la capacité d'agir autrement à t2 si :

- il est possible qu'il agisse autrement (compte tenu des lois de la nature et du passé actuel), et cette possibilité lui est accessible, ou si :
- il est *causalement* possible qu'il agisse autrement (compte tenu des lois de la nature et du passé actuel) et cette possibilité lui est accessible, ou si :
- aucun obstacle ne l'empêche d'agir autrement et il ne subit ni pression, ni contrainte, ni trouble psychique, physique ou moteur, etc. qui l'empêche d'agir autrement.

Fischer interprète son exemple en comprenant de la troisième façon les conditions de la capacité. En précisant que dans cet exemple, le déterminisme causal n'est pas vrai, il cherche à satisfaire les libertariens qui pensent de la deuxième façon, c'est-à-dire ceux qui pensent que la capacité suppose la possibilité causale. Mais pour un libertarien qui admet la première formulation, l'exemple de Fischer est peu convaincant.

Un libertarien peut facilement admettre que le vieux marin possède à midi la *capacité générale* d'aller en mer, car même s'il ne peut naviguer à midi, le vieux marin est en bonne santé et maîtrise toujours l'art de la navigation, de même qu'un pianiste est capable de jouer du piano même lorsqu'il n'a aucun instrument à sa disposition au moment où on lui attribue cette

¹⁸ J. M. Fischer, *The Metaphysics of Free Will*, p. 81.

capacité. Mais cette capacité générale d'agir autrement n'est pas suffisante pour la responsabilité morale. Pour un incompatibiliste, celle-ci requiert la capacité stricte d'agir autrement, qui suppose la possibilité causale, voire la possibilité *tout court* d'agir autrement. La première s'oppose au déterminisme causal, et la deuxième, au déterminisme en général.

Ainsi, la discussion aboutit à une impasse dialectique, comme le dit Fischer¹⁹ lui-même: le compatibiliste s'oppose à l'argument du transfert en présentant des contre-exemples trop peu convaincants pour un incompatibiliste. Celui-ci adhère donc toujours à l'argument du transfert. Pour le premier, le conditionnel et l'attribution de capacité peuvent être vrais en même temps ; pour le second, cela est impossible. Et aucun des deux ne peut justifier son jugement. Pour les incompatibilistes, la fixité des lois de la nature et du passé actuel sont des évidences premières. Aussi, en vertu du principe de Reid²⁰, ils peuvent déclarer que ces principes sont vrais si aucun contre-exemple n'oblige à les abandonner ou à les réviser. Les compatibilistes devront donc inventer une autre stratégie pour contrer l'argument du transfert.

Conclusion. Avant d'aborder les autres stratégies, faisons le point sur la critique de FL et de FP. Le compatibilisme peut procéder en corrigeant ou en rejetant FL, ou FP, ou ces deux principes à la fois. S'il veut modifier FL sans toucher à FP, il doit dire que nous sommes capables d'agir de telle sorte que certaines lois de la nature (du monde actuel) sont fausses. Cela revient à dire que nous avons accès à des mondes possibles où les lois de la nature ne sont pas les mêmes que dans le monde actuel. Pour éviter une idée aussi contre-intuitive, il peut choisir d'adopter FL et de réviser plutôt FP. Mais s'il révisé FP, il doit supposer que nous pouvons agir de telle sorte que le passé est très différent du passé actuel. En effet, puisque la thèse déterministe suppose une connexion étroite entre les événements successifs (en vertu des lois de la nature), si un événement du monde *wa* ne s'est pas produit dans le monde actuel, les lois de la nature restant exactement les mêmes, cette différence entre le passé dans *wa* et le passé dans le monde actuel implique d'autres différences à des moments antérieurs, et peut-

¹⁹ J. M. Fischer, *The Metaphysics of Free Will*, p. 83.

²⁰ Fischer lui-même n'invoque pas ce principe.

être même une série infinie de différences entre les événements du monde wa et ceux du monde actuel. Pour cette raison, un compatibilisme qui révisé FP sans réviser FL implique lui aussi que nous avons accès à des mondes possibles très différents du monde actuel.

Nous voyons ainsi que pour réviser un des principes de fixité, il faut réviser à la fois FL et FP, en soutenant que nous pouvons faire des actions impliquant un « miracle local », violation ponctuelle des lois de la nature dans la série des événements passés. Cette troisième thèse a un mérite que n'ont pas les deux autres : elle évite de supposer que nous ayons accès à des mondes possibles largement différents du monde actuel. Cependant, elle est peu vraisemblable pour les raisons que j'ai indiquées. Premièrement, elle suppose de réviser à la fois FP et FL, deux principes fortement intuitifs. Deuxièmement, comme l'a montré Helen Beebe, elle requiert une conception hybride des lois de la nature selon laquelle les lois de la nature imposent des limites plus strictes aux actions qu'aux autres événements.

En conclusion, les tentatives pour défendre une position compatibiliste en contestant FP ou FL ont des conséquences trop étranges. FP et FL sont des prémisses solides, et comme l'argument du transfert repose sur ces deux prémisses, la seule manière de le réfuter est de montrer qu'il n'est pas valide. Cette stratégie est celle que nous allons étudier maintenant.

3.4 L'argument du transfert est-il valide ?

L'argument de la conséquence repose, outre FP et FL qui constituent les prémisses, sur le principe du transfert d'impuissance :

(TI) Si Np et si $N(p \rightarrow q)$, alors Nq .

(Si personne ne peut empêcher que p , et si personne ne peut empêcher que $(p \rightarrow q)$, alors personne ne peut empêcher que q).

Or ce principe semble étroitement lié au principe du transfert de nécessité qui affirme :

(TN) Si $\Box p$ et si $\Box(p \rightarrow q)$, alors $\Box q$.

Dans un article qui a eu beaucoup d'influence²¹, Michael Slote a montré que le principe du transfert de nécessité, malgré son évidence apparente, n'est valable que pour certains genres de nécessité, comme la nécessité logique. Tout d'abord - le fait est connu -, ce principe ne vaut pas pour la nécessité *épistémique*. Ensuite, Slote expose un contre-exemple dans le domaine de la nécessité *déontique*. Puis, ayant montré que le principe du transfert n'est valable ni du point de vue épistémique ni du point de vue déontique, et ayant ainsi jeté le doute sur sa validité dans le domaine de la nécessité *causale ou métaphysique*, il propose un argument destiné à prouver qu'il ne vaut pas non plus dans ce dernier domaine. La conclusion de l'article est que le déterminisme est immunisé contre l'objection du transfert, car le déterminisme est une thèse relative à la nécessité *causale* et non à la nécessité *logique*. Avant d'examiner l'argumentation de Slote, rappelons la forme de l'argument qu'il critique :

1. NP
2. NL
3. N(P et L) (en vertu de 1 et 2)
4. N((P et L)→F). Donc :
5. NF (en vertu de 3 et de 4).

L'argument de la conséquence repose sur trois prémisses : les propositions 1 et 2 affirment la fixité du passé et des lois de la nature, et la troisième prémissse (la proposition 4) n'est autre qu'une formulation de la thèse déterministe, selon laquelle F, l'ensemble des événements présents et futurs, découle de manière inévitable de la conjonction des lois de la nature et des faits appartenant au passé actuel (le déterminisme admet donc que $N((P \text{ et } L) \rightarrow F)$). Nous avons vu que les propositions 1 et 2 sont difficilement contestables. Puisque la formulation de la thèse déterministe utilisée dans l'argument est généralement admise, la seule manière qui reste de réfuter l'argument est de contester sa validité, ce qui peut se faire de deux manières : en niant que l'on puisse valablement passer de 1 et 2 à 3, ou en niant que l'on puisse valablement passer de 3 et 4 à 5. Michel Slote a essayé de montrer que ces deux inférences, valides avec l'opérateur « $\square$ », deviennent douteuses avec l'opérateur « N » exprimant

²¹ Michael Slote, « Selective Necessity and the Free Will Problem », *Journal of Philosophy* 79 (1982), p. 5-24.

l'impuissance (« nul ne peut empêcher que »). Les deux principes suivants ne s'appliqueraient donc pas au connecteur « N » :

(Règle d'agglomérativité) : si $\Box p$, et si $\Box q$, alors $\Box(p \text{ et } q)$.

(Règle de clôture de l'implication) : si $\Box p$, et si $\Box(p \rightarrow q)$, alors $\Box q$.

Reprenons l'argumentation à partir du début. Slote montre d'abord que ces règles ne s'appliquent pas à toutes les formes de nécessité. Par exemple, « on admet généralement que « A sait que p » et « A sait que $(p \rightarrow q)$ » n'impliquent pas « A sait que q »... On peut manquer de faire une inférence qu'on est censé faire »²². Par exemple, si l'implication de p à q n'est pas directe mais suppose des calculs mathématiques, il est possible de savoir que p, et que p implique q, sans savoir que q. Le principe du transfert ne s'applique donc pas à la nécessité épistémique²³. car la deuxième règle (clôture sous l'implication) n'est pas valide pour cette forme de nécessité.

Il ne s'applique pas non plus à la nécessité déontique. Par exemple, ces trois propositions sont compatibles :

1. J'ai l'obligation d'aller à cette soirée.
2. J'ai l'obligation d'apporter du champagne si je vais à cette soirée.

Pourtant, il semble que

3. Je n'ai pas l'obligation d'apporter du champagne.

Il n'y a pas de règle de bienséance me commandant d'apporter du champagne, mais il y en a une ordonnant de faire ce qu'on a promis ; or j'ai promis d'y aller, et il était convenu que les hommes y participant devaient apporter du champagne. L'inférence n'est donc pas close

²² Michael Slote, « Selective Necessity and the Free Will Problem », p. 5.

²³ Pour une discussion détaillée de la question, cf. Fischer, *The Metaphysics of Free Will*, ch. II, p. 29-36.

sous l'implication, puisque (1) et (2) n'impliquent pas que j'ai l'obligation d'apporter du champagne. De même, on ne peut dire :

1. J'ai l'obligation d'aller chez Antoine dimanche prochain.
2. J'ai l'obligation de saluer sa mère si elle se trouve chez lui en même temps que moi.
3. Or elle sera chez lui dimanche prochain. Donc :
4. J'ai l'obligation de saluer sa mère dimanche prochain.

D'après Slote, la règle de clôture sous l'implication ne s'applique pas ici, car la nécessité déontique est *sélective* : sauf dans le cas d'obligations que nous avons envers tout être humain, elle est liée à des personnes ou à des circonstances particulières (on n'a pas les mêmes devoirs envers ses parents, ses amis, ses collègues et ses supérieurs hiérarchiques). Aussi est-il étrange de dire que j'ai l'obligation d'aller saluer la mère de mon ami Antoine. Mon obligation de tenir ma promesse envers Antoine en allant chez lui dimanche est sans rapport avec mon obligation de saluer sa mère si je la vois. Pour cette même raison, la nécessité déontique n'est pas *agglomérative*. On ne peut tenir un raisonnement comme :

1. J'ai le devoir d'aller chez Antoine dimanche.
2. J'ai le devoir de saluer sa mère si je la vois. Donc
3. J'ai le devoir de (aller chez Antoine dimanche et saluer sa mère).

Il est douteux que cette dernière obligation découle rigoureusement des deux précédentes. Ainsi, le caractère « sélectif » des modalités déontiques n'autorise pas le transfert que nous considérons comme parfaitement valide dans le domaine de la nécessité logique.

Mais qu'en est-il de la nécessité causale ? Le transfert sur lequel repose l'argument incompatibiliste est un transfert d'impuissance ou de nécessité causale. Slote estime qu'il s'agit d'une nécessité aléthique. Il doit donc montrer que la nécessité aléthique n'autorise pas l'usage des deux règles énoncées plus haut. La non-accidentalité, selon Slote, est un bon exemple de notion métaphysique assez proche de la notion exprimée par le connecteur « N », et qui n'est pas *agglomérative*.

En effet, on ne peut tenir le raisonnement suivant (où NA signifie « il n'est pas accidentel que »):

1. NA (Jim est à la banque à 10h).
2. NA (Jules est à la banque à 10h).
3. Donc NA (Jim et Jules sont à la banque à 10h).

Ce raisonnement n'est pas valide. Cela s'explique aisément par le fait que le hasard est la rencontre de deux chaînes causales indépendantes, comme l'a bien montré Aristote. Un homme peut rencontrer son débiteur alors qu'ils ne s'étaient pas donné rendez-vous et que chacun y venait avec une intention différente, l'un pour acheter quelque chose, et l'autre pour y rencontrer un ami. Dans ce cas, bien qu'aucun des deux ne s'y trouve par hasard à ce moment-là, il est accidentel qu'ils y soient en même temps, car personne n'a voulu et fait qu'ils y soient en même temps. Là encore, le caractère sélectif de la non-accidentalité fait qu'elle n'autorise pas le transfert.

Dans cet argument, Slote ne définit pas clairement la non-accidentalité, mais on peut penser que l'opérateur NA signifie « il fait partie du plan de quelqu'un que ». Imaginons le scénario suivant. Le patron de Jim l'a envoyé à la banque à 10h pour une opération bancaire tout à fait banale. Jules est un employé de cette banque. Tous les jours sauf le dimanche, il est à la banque à 10h. Dans ce cas, il n'est pas accidentel que Jules soit à la banque car cela est prévu dans son emploi du temps et correspond à la volonté de ses supérieurs. De même, le patron de Jim a voulu que Jim soit à la banque à 10h. Mais personne n'a eu l'intention que Jules et Jim soient à la banque au même moment. Le raisonnement exposé plus haut n'est donc pas valide, car NA est sélectif, comme le dit Slote.

Mais, précisément pour cette raison, on peut essayer de reformuler le raisonnement en indexant l'opérateur NA. L'opérateur N utilisé dans le transfert d'impuissance peut être indexé sur un agent (ou un groupe d'agents), ou ne pas être indexé, auquel cas il signifie « nul ne peut empêcher que ».

De même, l'opérateur NA peut être indexé ou non, ce qui laisse le choix entre les deux formules suivantes :

(A) $N_s A p_t, N_s A q_t$, donc $NA_s (p_t \wedge q_t)$.

(B) $NA p, NA q$, donc $NA(p \wedge q)$.

Nous avons vu que le deuxième raisonnement n'est pas valide (c'est ce que montre l'exemple de Slote). Le raisonnement A, en revanche, semble correct. Par exemple :

1. Si $NA_{Paul}(\text{Jules est à la banque à 10h})$.
2. et $NA_{Paul}(\text{Jim est à la banque à 10h})$.
3. alors $NA_{Paul}(\text{Jules et Jim sont à la banque à 10h})$.

Si NA_s veut dire « Il y a un plan de S qui prévoit que », ce raisonnement est douteux, car même s'il est conforme aux intentions de Paul que Jules soit à la banque à 10h et que Jim y soit à 10h, ces deux intentions peuvent appartenir à deux projets différents. En outre, il peut y avoir des incohérences dans les intentions ou les désirs d'une personne. Cela est particulièrement clair dans le cas des désirs. Par exemple, de

1. « Paul désire manger des lasagnes à t », et de
2. « Paul désire manger une pizza scandinave à t », on ne peut conclure :
3. « Paul désire manger des lasagnes et une pizza scandinave à t ».

Ces exemples suggèrent que si l'opérateur NA n'est pas agglomératif, c'est probablement parce qu'il introduit des contextes obliques. De « Paul croit que p » et « Paul croit que $p \rightarrow q$ », on ne peut déduire « Paul croit que q », parce qu'on ne peut présupposer une parfaite cohérence dans les croyances de Paul. De même, de « Paul désire que p » et « Paul désire que q », il ne s'ensuit pas que Paul désire que « $p \wedge q$ », car il peut avoir des désirs incompatibles entre eux.

De même enfin, puisque le connecteur NA, indexé sur Paul, signifie « Il appartient aux intentions de Paul que », il n'est pas étonnant que NA ne soit pas agglomératif. D'ailleurs, le problème persiste lorsque NA n'est pas indexé et signifie « Il appartient aux intentions de tous

que ». Mais ce défaut n'existe pas dans le transfert d'impuissance, car le connecteur « N » ne se réfère pas aux désirs ou aux pensées de quelqu'un. C'est pourquoi le fait que N puisse être indexé, comme NA, ne prouve nullement qu'il soit sélectif et pose les mêmes problèmes que NA. La différence entre ces deux notions est considérable. Slote écrit : « on peut difficilement nier que la non-accidentalité soit une notion causale/nomologique portant sur la manière dont les choses sont dans le monde, et une forme de nécessité *aléthique* en particulier. Pour ceux qui distinguent les vérités universelles du genre des lois, c'est-à-dire les lois de la nature, et celles qui sont d'un autre genre que les lois, nous pouvons qualifier celles-ci d'accidentelles. Et il est naturel d'identifier les accidentelles avec les contingentes, et de traiter la non-contingence comme une forme de nécessité »²⁴. O'Connor répond que ceux qui qualifient les vérités universelles d'un autre genre que les lois d'accidentelles n'emploient sans doute pas ce terme « dans le sens plutôt mal défini que Slote a à l'esprit ». Je pense aussi qu'on peut difficilement considérer comme voisins des connecteurs aussi différents que :

- « Il appartient à un plan de quelqu'un que »
- « Personne ne peut empêcher que »

Le premier relève de la généralisation existentielle alors que le second exprime une généralisation universelle. Donc même si ces deux connecteurs exprimaient chacun une nécessité aléthique, ils sont trop différents pour supposer que ce qui est vrai de l'un est aussi vrai de l'autre. Ces remarques s'appliquent également à la clôture sous l'implication.

Reprenons l'exemple de Slote :

1. NA (Je suis à la banque à 10h).
2. NA (Je suis à la banque à 10h $\rightarrow$ Je suis en vie à 10h).

Mais il est faux que :

3. NA (Je suis en vie à 10h).

²⁴ Michael Slote, « Selective Necessity and the Free Will Problem », p. 15-16.

Le fait que je sois en vie à 10h peut ne pas faire partie du plan de mon supérieur, c'est pourquoi, selon Slote, (3) ne découle pas rigoureusement des prémisses (1) et (2). Mais là aussi, le fait que la non-accidentalité ne se transfère pas vient du fait que NA renvoie aux intentions d'un agent, ce qui introduit l'*oratio obliqua*. De « S croit que p » et « p implique q », on ne peut déduire « S croit que q », et de la même façon, de « S veut que p » et « p implique q », on ne peut déduire « S veut que q ». Mais le connecteur « N », au contraire, ne renvoie pas aux désirs ou aux intentions d'un agent. De plus, comme le note Timothy O'Connor, si le plan de mon supérieur n'implique pas le fait que je sois en vie à 10h, comme le dit Slote, alors la prémisse (2) est fausse, ce qui suffit à expliquer la fausseté de la conclusion. Ainsi, le contre-exemple de Slote ne porte pas contre une version du principe du transfert de non-accidentalité dans laquelle le connecteur NA apparaît au début de chaque prémisse. Elle réfute seulement le transfert suivant :

1. NAp
2. $p \rightarrow q$
3. NAq

où NA n'apparaît pas dans la deuxième prémisse. Enfin, nous avons vu que la réfutation du transfert de non-accidentalité élaborée par Slote suppose que NA soit défini en référence aux croyances et aux intentions conscientes de l'agent. Or pour que NA exprime une nécessité aléthique, il faut qu'il exprime une nécessité objective, indépendante de ce que pense ou désire l'agent de façon consciente. Et dans ce cas, si NAp et si $NA(p \rightarrow q)$, alors NAq.

Cette modification du sens de NA ne le rend pas agglomératif, mais en faisant cette remarque sur le sens de NA supposé dans les exemples de Slote, O'Connor montre que ces exemples ont peu de rapport avec le transfert d'impuissance. Même améliorés, ces contre-exemples reposent sur le fait que NA introduit l'*oratio obliqua*. Sur ce point, la non-accidentalité dont parle Slote ressemble d'avantage à la nécessité épistémique qu'à une nécessité objective comme celle du transfert d'impuissance. Slote n'a donc pas réfuté le transfert d'impuissance.

3.5 L'argument du transfert prouve-t-il l'incompatibilité du déterminisme et de la responsabilité morale ?

Nous avons vu que l'argument du transfert repose sur des prémisses solides et que les objections de Slote contre la validité du transfert ne sont pas recevables. Ainsi, puisque cet argument est valide et fondé sur des prémisses vraies (FL et FP), sa conclusion devrait être bonne. Il y a cependant une dernière stratégie compatibiliste pour neutraliser l'argument du transfert. En effet, un compatibiliste peut finalement admettre que cet argument part de prémisses vraies, est valide, et aboutit à une conclusion juste, tout en déclarant que cette conclusion est sans véritable portée lorsqu'il s'agit de savoir si le déterminisme est compatible avec la responsabilité morale. Rappelons que dans l'argument du transfert sous sa forme la plus répandue, l'opérateur « N » n'exprime pas directement la non-responsabilité morale, mais l'impuissance (ou la « *power-necessity* »), notion que l'on suppose plus proche de la nécessité absolue exprimée par le connecteur « $\square$ ». Il s'ensuit qu'un compatibiliste peut admettre Nsq tout en soutenant que S est moralement responsable de q. De même, puisqu'il y a différentes sortes de capacités, il peut admettre que le principe du transfert est valable pour certaines sortes de capacités, mais pas pour celles dont dépend la responsabilité morale. Cette position est celle de Fischer²⁵. Avant de l'examiner, je vais d'abord discuter une position suggérée par Slote à la fin de son article²⁶. Le point commun entre ces deux positions est qu'elles font appel à l'intuition en déclarant qu'il y a des cas où même si le déterminisme est vrai (ou même si Nsp et Ns(p→q)), S semble bien responsable de q. La force intuitive du jugement « S est responsable de q » serait plus grande que la force intuitive du principe TI (transfert d'incapacité).

Michael Slote n'a pas réussi à prouver la non-validité du transfert d'impuissance, mais on ne peut pas prouver non plus sa validité. Il reste donc possible, pour le compatibiliste, de *supposer* ce que Slote voulait prouver, à savoir que S est responsable de q et que le transfert

²⁵ J. M. Fischer, *The Metaphysics of Free Will*, p.92, et J. M. Fischer & M. Ravizza, *Responsibility and Control*, p.168. Je discuterai leur argument contre le principe du transfert de non-responsabilité dans la section 3.7.

²⁶ Michael Slote, « Selective Necessity and the Free Will Problem », *Journal of Philosophy* 79 (1982), p. 5-24. Repris dans Fischer (2005a). Le passage auquel je me réfère ici se trouve p. 122-123 dans Fischer (2005a), vol.2.

d'impuissance n'est pas valide : puisque certaines formes de nécessité ne sont pas agglomératives (ou pas closes sous l'implication), il est possible que N ne le soit pas non plus.

Slote pense que dans l'argument de la conséquence, les prémisses sont vraies parce que les lois de la nature et les faits passés antérieurs à la naissance de l'agent sont indépendants de ses désirs et de ses intentions. Nsp signifie alors « p est indépendant des désirs et des intentions de S ». Dans ce cas, l'argument de la conséquence n'est pas valide, car puisque l'événement q dont il s'agit dans la conclusion est soit une action de l'agent, soit l'effet d'une de ses actions, q *n'est pas* indépendant des désirs ou des intentions de l'agent. Ceci explique, selon Slote, pourquoi l'argument du transfert n'est pas valide. En effet, il est manifeste que ces trois propositions sont compatibles :

- p est indépendant des désirs et des intentions de S.
- $(p \rightarrow q)$ est indépendant des désirs et des intentions de S.
- q dépend (ou résulte) d'un désir (ou d'une intention) de S.

Or cette interprétation de l'argument du transfert est typiquement compatibiliste. Pour les libertariens qui défendent cet argument, Nsp ne signifie pas que p ne dépend pas de S mais qu'il n'en dépend pas *ultimement*. Si l'agent n'est pas la source ultime (ou au moins *une* source ultime) de p, ou de $p \rightarrow q$, il ne peut être une source ultime de q. C'est pourquoi il n'en a pas la *responsabilité ultime*. L'explication donnée par Slote est donc trop éloignée du vrai sens de l'argument de la conséquence. Elle ne donne aucune raison de rejeter cet argument si Nsp veut dire « S n'est pas une source ultime de p » ou « S n'exerce pas de contrôle ultime sur p ».

En s'inspirant de Slote, un compatibiliste peut seulement *supposer* que l'argument de la conséquence n'est pas valide compte tenu du fait qu'un raisonnement de même forme n'est pas valide si l'on substitue à « N » un opérateur exprimant la nécessité épistémique, la nécessité déontique ou la non-accidentalité. Mais ces trois nécessités sont très éloignées de l'opérateur « N » : elles ne sont pas des nécessités aléthiques, mais sont toutes les trois relatives aux pensées de certains individus. On ne peut donc pas déclarer, sur la base de ces seules raisons, que la force intuitive de l'argument incompatibiliste est une apparence trompeuse.

La position de Fischer, quant à elle, consiste à ne pas nier la validité de l'argument du transfert, mais à en limiter la portée. Il reconnaît la force intuitive de l'argument mais veut montrer que le contrôle ultime dont il s'agit dans cet argument n'est pas nécessaire pour la responsabilité morale. C'est la dernière stratégie possible pour neutraliser l'argument du transfert. Elle est difficile à critiquer car elle ne contredit pas nos intuitions de manière aussi flagrante que les théories fondées sur la remise en cause des principes de fixité.

Comme je l'ai dit plus haut, Fischer défend l'argument du transfert tout en soutenant qu'il ne concerne pas la responsabilité morale. Pour montrer sa solidité, il le formule de manière à le rendre indépendant de tel ou tel principe modal particulier. Cette version simple et intuitive de l'argument échappe ainsi aux objections de Slote, par exemple. Fischer²⁷ pose d'abord les principes de fixité :

(FP) Pour tout agent S, toute action A et tout instant t, si (si S faisait A à t, alors un des faits passés relativement à t n'aurait pas été un fait), alors S ne peut pas, à t, faire A.

(FL) Pour tout agent S, toute action A et tout instant t, si (si S faisait A à t, alors une loi de la nature qui vaut actuellement ne vaudrait pas) alors S ne peut, à t, faire A.

Le déterminisme est la thèse selon laquelle une description complète de l'état du monde à t, jointe à une formulation complète des lois de la nature, implique toutes les vérités sur des événements postérieurs à t. Donc si le déterminisme est vrai et si E1 est l'état complet du monde à t1, un des conditionnels suivants doit être vrai :

1. Si S se retenait de faire A à t2, E1 n'aurait pas été l'état complet du monde à t1.
2. Si S se retenait de faire A à t2, alors une loi naturelle qui vaut actuellement ne vaudrait pas.
3. Si S se retenait de faire A à t2, alors soit E1 n'aurait pas été l'état complet du monde à t1, soit une loi naturelle qui vaut actuellement ne vaudrait pas.

²⁷ J. M. Fischer, *The Metaphysics of Free Will*, p. 62.

Puisque FP et FL sont vrais et que chacun des conditionnels contredit au moins FP ou FL, le déterminisme est incompatible avec la possibilité d'agir autrement (en l'occurrence, il est incompatible avec la possibilité pour S de ne pas faire A).

Comme le dit Fischer²⁸, la force intuitive de cet argument reste grande même si l'on peut réfuter telle ou telle version de l'argument dans laquelle on utilise un principe modal plus ou moins analogue à TN, le transfert de nécessité logique. Or il est a priori très difficile de réfuter tous les principes modaux sur lesquels peut se fonder l'argument du transfert, et il n'est même pas certain que cet argument dépende nécessairement d'un tel principe. Fischer propose finalement une dernière version de l'argument incompatibiliste²⁹ qui ne dépend explicitement ni de TI, ni d'une formulation particulière des principes de fixité. Elle repose sur l'idée qu'un agent « peut faire A seulement si son action de faire A peut être une extension du passé actuel, les lois étant fixes »³⁰. Supposons que S bouge sa jambe à t2. Le déterminisme suppose que l'état du monde à t1 implique, en vertu des lois de la nature, que S bouge sa jambe à t2. Or un agent ne peut accomplir une action que si cette action peut être une extension du passé actuel, les lois demeurant fixes. Puisque l'action de S de maintenir sa jambe immobile à t1 ne peut pas être une extension du passé actuel (les lois demeurant fixes), il s'ensuit que S ne peut pas maintenir sa jambe immobile à t2.

Ainsi, le déterminisme implique que pour toute action A et tout instant t, je n'aurais pas pu faire une autre action, et il y a un sens du mot « capable » dans lequel on peut dire que je n'étais pas capable d'agir autrement. La difficulté qui subsiste est de savoir quelle sorte de capacité suppose la responsabilité morale : est-ce bien celle-là ou une autre ? Appelons cette capacité la *capacité stricte*. La capacité stricte (de S) de faire A à t suppose que l'événement « S fait A à t » ne soit impossible *en aucun sens du terme*, c'est-à-dire qu'elle ne soit ni impossible logiquement, ni impossible en vertu des lois de la nature, ni impossible en vertu des lois de la nature et du passé actuel. Selon Fischer, la responsabilité ne requiert pas cette sorte de capacité, même si elle suppose que l'agent ait un certain contrôle de ses actions.

²⁸ *Ibid.*, p. 64.

²⁹ C'est la « version basique » de l'argument incompatibiliste. Cf. Fischer, *The Metaphysics of Free Will*, p. 88.

³⁰ *Ibid.* p. 88.

La position de Fischer et toute position du même genre a besoin, pour répondre à l'argument du transfert de façon satisfaisante, de présenter une conception intelligible et suffisamment intuitive des capacités requises par la responsabilité morale³¹. En outre, elle a besoin d'une réfutation de TNR, le transfert de non-responsabilité³². En effet, si TNR est réfuté par des contre-exemples, il est possible de soutenir que le transfert d'impuissance (TI) et les arguments moins formels utilisés par Fischer pour exprimer la même idée ne justifient pas TNR et que l'impuissance dont il s'agit dans TI n'exclut pas la responsabilité morale. Mais si TNR n'est pas réfuté, l'hypothèse la plus probable est que TI est vrai et que sa vérité explique celle de TNR.

Plus loin³³, je défendrai TNR sous plusieurs formes. Je soutiendrai que TNR est au moins aussi solide que TI. L'argument fondé sur TNR a une grande force intuitive et est l'argument direct par excellence, car il se réfère directement à la notion de responsabilité morale (non à la capacité, à l'incapacité, ou à la nécessité causale). Mais cela n'enlève pas tout intérêt aux arguments moins directs fondés sur le transfert d'impuissance. Celui-ci peut renforcer le principe de non-responsabilité ou lui conférer une certaine probabilité a priori. Inversement, il est probable a priori que TNR suppose la validité d'au moins une version de TI, puisque la notion de responsabilité paraît liée à la notion de pouvoir ou de capacité.

Avant d'examiner les arguments utilisés contre TNR, j'essaierai, dans la prochaine section, de montrer que TI peut être dérivé d'une version corrigée de TC, le transfert de capacité. Le lecteur pourra passer directement, s'il le souhaite, à la section 3.7, car cette démonstration n'est pas nécessaire pour justifier TI, finalement plus intuitif et plus facile à défendre que TC. Mais comme on peut supposer qu'il y a un lien conceptuel entre TC et TI, il est tentant de démontrer TI à partir d'une version du transfert de capacité.

³¹ Tâche entreprise par Fischer dans *The Metaphysics of Free Will*, ch. 7 et 8, puis par Fischer et Ravizza dans *Responsibility and Control*.

³² Cf. Fischer & Ravizza, *Responsibility and Control*, ch.6, p. 154 et s., où Fischer et Ravizza exposent des contre-exemples à TNR.

³³ Dans les sections 3.7 et 3.8.

3.6 La stratégie diodorienne

Dans les sections précédentes, nous avons examiné les différentes solutions compatibilistes pour critiquer l'argument de la conséquence fondé sur le transfert d'impuissance (TI). Elles se répartissent en trois catégories : la première catégorie rejette une des prémisses de l'argument, c'est-à-dire un des principes de fixité, en le remplaçant par un autre principe moins strict. Ce type de solution fait référence à des mondes possibles éloignés du monde actuel ou à des capacités non-causales, or ces deux solutions semblent peu cohérentes. La deuxième solution est d'accepter les prémisses de l'argument en contestant sa validité. L'argument n'est pas valide si la modalité exprimant l'impuissance (l'opérateur « N ») n'est pas agglomérative ou si elle ne permet pas la clôture sous l'implication. Slote a donné des exemples de modalités qui ont ce défaut alors qu'elles expriment une sorte de nécessité, mais ces modalités sont très éloignées de l'impuissance (ou « *power-necessity* ») à cause de leur composante épistémique. En effet, elles font toutes référence aux croyances ou aux désirs de certains individus. L'objection de Slote ne permet donc pas de neutraliser l'argument du transfert. La dernière solution compatibiliste est de reconnaître à la fois la validité et les prémisses de l'argument du transfert, donc aussi sa conclusion, mais en montrant que cette conclusion est sans rapport avec la notion de responsabilité : la conclusion est que l'agent est incapable d'empêcher que q, mais cela ne prouverait pas qu'il ne soit pas responsable de q. Défendre cette thèse revient à admettre TI tout en rejetant TNR, bien que ces deux principes semblent étroitement apparentés. Cette thèse a donc besoin d'un argument réfutant directement TNR. Avant de discuter le principe TNR, j'essaierai ici de répondre à un argument qui vise à fragiliser TI en montrant qu'il n'est pas possible de le dériver de TC, le principe du transfert de capacité. Cet argument a été exposé par Kapitan³⁴. Dans cette section, je reprendrai donc une partie de son raisonnement et de ses exemples. J'épargnerai au lecteur le détail de son argumentation en m'appuyant d'avantage sur des contre-exemples. À la fin, je répondrai à son objection en proposant une version corrigée de TI qui peut être dérivée d'une version correspondante de TC.

³⁴ Tomis Kapitan, "A Master Argument for Incompatibilism ?" (2002). In Kane (2002b).

Les modalités logiques se définissent les unes à partir des autres. Il est tentant de penser que les connecteurs de capacité et d'incapacité peuvent aussi se définir mutuellement. Cela permettrait de dériver TI de TC, et ainsi, de renforcer la plausibilité de TI. La recherche d'une formulation satisfaisante de TC montre que ce principe n'est pas plus évident que TI, mais *moins* évident. Je pense donc qu'en dérivant TI de TC, on ne tire pas TI d'un principe plus évident et plus élémentaire, mais on établit des connections entre deux principes de manière à les renforcer tous les deux, la probabilité d'une idée étant augmentée lorsqu'elle est rattachée à d'autres idées qui forment un système cohérent et résistant aux critiques.

TI et TC peuvent prendre plusieurs formes à cause des différents sens du mot « capacité ». Nous pouvons distinguer la capacité large, la capacité causale et la capacité stricte. David Lewis³⁵ se réfère à une capacité large en soutenant qu'on peut accomplir une action dont la réalisation implique la violation d'une loi de la nature à un moment antérieur. En effet, cette capacité n'est pas celle de *provoquer* un événement qui viole une loi. Lewis suppose ainsi une forme de capacité qui n'est pas de nature causale. Cette notion est problématique car ce que nous appelons normalement une capacité est une sorte de contrôle causal, et pas seulement la possibilité de faire un acte *X impliquant* un événement *Y* à un instant *t*. Si j'ai la capacité de faire *X*, mais que *Y* ne peut pas être produit par moi, je n'ai pas de capacité proprement dite de faire-un-acte-*X*-tel-qu'un-événement-*Y*-se-produit-à-*t*, mais *Y* est hors de portée de mes capacités. De plus, les compatibilistes qui rejettent FL et FP admettent des versions causales de ces deux principes : FLc et FPc. Il est donc possible de contrer leurs objections en utilisant une version explicitement causale de l'argument du transfert. On remplace alors le connecteur « N » par un connecteur d'impuissance causale « Nc », et on raisonne de la même façon qu'avec N : Si le déterminisme est vrai, alors (P et L) → F. Compte tenu de la fixité du passé et des lois de la nature, le déterminisme autorise donc le raisonnement suivant (où la proposition (4) exprime la thèse déterministe) :

1. NcP
2. NcL
3. Nc(P∧L) en vertu de (1) et (2)

³⁵ David Lewis, "Are we free to break the laws ?" (1981), in Watson (2003). Cf. *supra*, p. 113 et s.

4. $\text{Nc}((P \wedge L) \rightarrow F)$. Donc :
5. $\text{Nc}F$ en vertu de (3) et de (4).

Puisque cette conclusion est fausse, le déterminisme doit être faux (c'est-à-dire la proposition (4)). La limite de cet argument vient du fait que pour un compatibiliste, la conclusion peut être fausse pour une autre raison : parce que les règles de clôture qui valent pour la nécessité logique ne valent pas pour la modalité utilisée ici. L'argument du transfert serait valide lorsqu'il porte sur la capacité large, mais ses prémisses (FL ou FP) seraient fausses, et la version causale de l'argument (portant sur la capacité causale) aurait le défaut inverse, car ses prémisses seraient vraies, mais sa validité serait douteuse.

En fait, même la capacité causale ne semble pas assez riche pour rendre compte des attributions de responsabilité. En effet, je peux être incapable *au sens strict* de faire ce que j'ai la capacité causale de faire (telle que nous l'avons définie). Par exemple, une fillette de cinq ans fait un *strike* la première fois qu'elle lance la balle au bowling. Elle est incapable de réitérer cette action, car elle n'a pas l'habileté requise pour cela. L'habileté diffère de la chance par la régularité, c'est-à-dire par un pourcentage de réussites relativement élevé. La capacité stricte est donc une capacité causale régulière.

Un compatibiliste peut la définir ainsi :

(Capacité stricte) S est capable au sens strict, à t, de faire que p si et seulement si il y a une action A telle qu'à t :

1. S est causalement capable de faire A, et
2. p est une conséquence explicable à partir de l'action de S lorsqu'il fait A.

Par « conséquence explicable », il faut comprendre une conséquence prévue et voulue, et obtenue non par chance, mais grâce à une compétence ou une habileté. Mais cette définition n'est pas encore complète. Par exemple, on peut avoir la capacité causale régulière de taper sur un clavier d'ordinateur « F1, K, A, N, O, N » puis « Enter » sans savoir que c'est le code secret permettant de résoudre un grave problème informatique qui est en train de détruire tout le contenu de l'ordinateur. Pour avoir la capacité stricte de taper le bon code, il faut avoir la

capacité causale régulière de taper ce code, mais aussi savoir que c'est le bon code. La capacité stricte inclut donc des conditions épistémiques. Il n'est pas nécessaire pour nous de posséder maintenant une définition plus précise de cette capacité. Les remarques précédentes permettent d'affirmer :

$$A^s \rightarrow A^c \rightarrow A^b$$

où A^s , A^c et A^b sont les opérateurs désignant respectivement la capacité stricte, la capacité causale, et la capacité large. On en déduit :

$$\sim A^b \rightarrow \sim A^c \rightarrow \sim A^s$$

ce qui revient à écrire :

$$N^b \rightarrow N^c \rightarrow N^s$$

(ce qui est inévitable au sens large l'est aussi au sens causal et au sens strict).

Ainsi, pour montrer que ce qui est causalement déterminé est inévitable au sens strict, il suffit de montrer qu'il l'est en un de ces sens.

Le fait que les modalités logiques et pratiques se définissent les unes à partir des autres suggère que le principe du transfert d'incapacité (TI) peut être dérivé d'un autre principe analogue concernant la capacité (TC). Ce raisonnement se fonde sur la similitude entre :

P1. Ap, q est une conséquence de p, donc Aq

et le principe diodorien :

M2. Tout ce qui est une conséquence d'une possibilité est lui-même possible.

et à partir de la similitude entre :

M1. Tout ce qui est une conséquence nécessaire d'une nécessité est lui-même nécessaire.

et :

P2. Np, q est une conséquence de p, donc Nq.

L'expression « est une conséquence de » peut s'interpréter de plusieurs manières. Elle peut vouloir dire « est un effet de » ou « est impliqué par ». Retenons ce dernier sens et remplaçons P1 par :

P3. $A p, p \rightarrow q$, donc $A q$.

Si on suppose que l'opérateur A fonctionne comme l'opérateur $\Diamond$ exprimant la possibilité logique, on peut déduire de P3 la proposition :

P4. $A \sim q, p \rightarrow q$, donc $A \sim p$.

Ensuite, on déduira :

P.5 $\sim A \sim p, p \rightarrow q$, donc $\sim A \sim q$.

Supposons en effet que $\sim A \sim p$ et que $p \rightarrow q$. Si $\sim A \sim q$ était fausse, nous aurions $A \sim q$. Or en vertu de P4, cette dernière proposition impliquerait $A \sim p$, ce qui contredirait la supposition. Or P5 équivaut à :

P6. $N p, p \rightarrow q$, donc $N q$.

Voilà en quoi consiste ce que Kapitan appelle la « stratégie diodorienne ». Pour défendre l'argument du transfert. La question qui se pose maintenant est-celle-ci : peut-on appliquer ce schéma de raisonnement à l'une ou l'autre des capacités que nous avons distinguées plus haut?

Avant de proposer une version de TC permettant de dériver TI, nous pouvons exclure plusieurs formulations de ces principes à l'aide de contre-exemples. Je commencerai par les versions les plus larges, pour m'intéresser ensuite à des versions plus strictes ou plus explicitement causales. La version la plus large de TI serait celle-ci :

(TLI) $N^b p, p \rightarrow q$, donc $N^b q$.

Puisque l'incapacité large de faire q implique l'incapacité causale et l'incapacité stricte de faire q (comme nous l'avons vu plus haut), si TIL est vrai, on peut en conclure $N^s q$. Le

raisonnement semble correct. Mais il ne réfute pas le compatibiliste. Reprenons l'exemple du vieux loup de mer utilisé par Fischer. Le service météo lui dit à 9h qu'il va faire mauvais à midi. Or si le service météo lui dit à 9h qu'il va faire mauvais à midi, il ne sort pas en mer à midi. Pourtant, à midi, il est capable de sortir en mer. Fischer pense qu'il est *strictement* capable de le faire. Le compatibiliste est en droit de nier la conclusion puis de chercher pourquoi elle est fausse. Il admet donc :

(a) $A^s \sim q$

qui est le « nerf » de la réponse compatibiliste.

Puis il est obligé d'en conclure qu'il faut rejeter

(b) $N^s q$ (q est strictement inévitable)

et donc aussi

(c) $N^b q$ (q est inévitable au sens large)

et puisque (P et L) implique q, alors C2 et C3 obligent le compatibiliste à nier

(d) $N^b(P \text{ et } L)$

Le recours aux principes de fixité n'est d'aucune utilité pour défendre (d), car tant que nous parlons de capacités *larges*, comme nous le faisons ici, les versions faibles de FP et de FL sont les seules pertinentes. Le passé et les lois ne sont pas donc totalement « fixes ». Tant que nous parlons de capacités larges, le vieux loup de mer est capable *au sens large* d'agir de manière à rendre q fausse, et puisque (P et L) implique q, il est capable de falsifier (P et L). Ainsi, il est possible de rejeter (d).

Qu'en est-il si nous parlons de capacités causales ? Dans ce cas, les principes de fixité sont incontestablement vrais. Les compatibilistes eux-mêmes admettent les versions causales de FP et FL. Nous pouvons donc écrire :

(e) $N^c(P \text{ et } L)$

(f) $N^s(P \text{ et } L)$

Voyons d'abord si certaines versions causales de TC et de TI sont réfutables par des contre-exemples. La version causale la plus simple de TI est :

(TCI 1) $N^c p, p \rightarrow q$, donc $N^c q$

Ce principe semble vrai. Mais on ne peut le dériver d'une version correspondante de TC :

(TCC 1) $A^c p, p \rightarrow q$, donc $A^c q$.

Cette version est incorrecte car si q est une vérité nécessaire (mathématique, par exemple) ou bien un fait largement antérieur à p , voire antérieur à la naissance de l'agent, le fait que l'agent puisse causer p n'implique pas qu'il puisse causer q . Ce problème vient du fait que la relation d'implication, dans le second membre de TCC1, a un sens trop vague : elle peut aller du présent vers le passé, ou d'un effet vers sa cause.

Le même problème se pose avec l'opérateur inventé par McKay et Johnson. Ils proposent d'utiliser A^w qui signifie « quelqu'un est capable de faire quelque chose pouvant aboutir à p ». Il est clair qu'en ce sens,

$N^w(P \text{ et } L)$

O'Connor, s'inspirant du travail de McKay et Johnson, propose cette version de TI :

(TCI 2) $N^w p, p \rightarrow q$, donc $N^w q$.

McKay et Johnson proposent celle-ci :

(TCI 3) $N^w p, N^w(p \rightarrow q)$, donc $N^w q$.

Ces principes semblent vrais, mais on ne peut les déduire des transferts de capacité correspondants, car ils sont faux. Les transferts de capacité correspondant à TCI 2 et à TCI 3 sont :

(TCC 2) $A^w p, p \rightarrow q$, donc $A^w q$. et :

(TCC 3) $A^w p, A^w(p \rightarrow q)$, $A^w q$.

Voici un cas où il semblent faux : « p » signifie « je crie « Diantre ! » » ; et q signifie « l'équipe de football du Ghana marque un point contre l'équipe de France. Supposons que je crie “Diantre !” seulement lorsque l'équipe du Ghana marque un but contre l'équipe de France. Dans ce cas, il peut arriver ceci : je suis capable de crier « Diantre ! » ; je suis capable de ne crier cela que lorsque l'équipe du Ghana marque un point contre l'équipe de France (et ainsi, je peux faire en sorte que p implique q) ; et pourtant je ne suis pas capable de faire en sorte que l'équipe du Ghana marque un point contre l'équipe de France, car cela ne dépend pas de moi. Ce contre-exemple réfute les trois versions de TCC. Il est donc impossible de dériver les versions de TCI que je viens de présenter en partant des transferts de capacité qui leur correspondent.

Pour corriger ce défaut, on peut remplacer le second membre par une proposition exprimant clairement une relation causale. Ainsi, q ne pourra plus être un événement antérieur à p, ni une cause ou une condition de p. Par exemple, nous pouvons partir de

(TCC 4) $A^c p$, (si p, p cause q), donc $A^c q$

où la relation d'implication, dans la deuxième proposition, est remplacée par ce qu'on peut appeler l'implication *causale*. Ensuite, tout comme on passe de P3 à P5, on peut passer de TCC 4 à

(2) $A^c \sim q$, (si p, p cause q), donc $A^c \sim p$

(3) $\sim A^c \sim p$, (si p, p cause q), donc $\sim A^c \sim q$

ce qui revient à

(TCI 4) $N^c p$, (si p, p cause q), donc $N^c q$

Or la proposition (2) est fausse. Soient p « Gunnar tire sur Ridley à t1 », et q : « Ridley meurt à t2 ». Si p, p cause q. Supposons que Gunnar tire effectivement. Dans ce cas, si je suis capable d'empêcher que Ridley meure à t2, je suis capable d'empêcher p (« Gunnar tire sur Ridley à t1 »). Seulement, (2) ne dit pas si p a lieu effectivement. Supposons que je sois médecin. Gunnar ne tire pas sur Ridley, mais celui-ci est atteint d'une maladie grave. En

soignant Ridley, je le sauve et il continue à vivre après t_2 . Si (2) est fausse, c'est parce que dans (2), aucun des deux membres de l'antécédent n'affirme p . Il est donc assez naturel de corriger TCC 4 en le remplaçant par :

(TCC 5) $A^c p$, (p et si p , p cause q), donc $A^c q$.

On en déduit :

(2) $A^c \sim q$, (p et si p , p cause q), donc $A^c \sim p$.

De là, on peut passer à :

(3) $\sim A^c \sim p$, (p et si p , p cause q), donc $\sim A^c \sim q$.

Supposons en effet que $\sim A^c \sim p$ et que (p et si p , p cause q). Cette dernière proposition implique $p \rightarrow q$. Nous avons donc : $\sim A^c \sim p$ et $p \rightarrow q$. A partir de là, nous pouvons raisonner ainsi : si $\sim A^c \sim q$ était fausse, nous aurions $A^c \sim q$. Or en vertu de (2), cette dernière proposition impliquerait $A^c \sim p$, ce qui contredirait la supposition. (3) est donc correcte. Or elle équivaut à :

(TCI 5) $N^c p$, (p et si p , p cause q), donc $N^c q$.

(TCI 5) est donc une formulation de TI qui peut être dérivée d'une formulation correspondante de TC. Désormais, j'appellerai ce principe le *transfert causal d'incapacité* (TCI), en laissant de côté toutes les versions précédentes, et de la même façon, je remplacerai la notation TCC 5 par TCC. Un compatibiliste pourrait objecter que vu le nombre des formulations de TC ou de TI qui peuvent être réfutées par des contre-exemples, TCI 5 ne peut être utilisé dans un argument contre la thèse compatibiliste. Je pense qu'il y a deux réponses à cette objection. Premièrement, nous pouvons remarquer qu'il est plus facile de trouver des exemples pour réfuter les différentes versions de TC que pour réfuter les différentes versions de TI. Le transfert d'incapacité a donc une plus grande évidence que le transfert de capacité. C'est la raison pour laquelle je ne pense pas qu'en dérivant TI de TC, on augmente beaucoup sa crédibilité. Mais c'est aussi une raison de penser qu'on ne peut pas jeter le doute sur TI en réfutant des versions de TC ou en montrant qu'on ne peut pas dériver TI de TC. Les versions de TI proposées par McKay et Johnson ainsi que O'Connor, ont en elles-mêmes une grande

force intuitive et ne paraissent pas réfutables, même si les versions correspondantes de TC le sont, comme je l'ai montré plus haut. Je considère donc ces deux principes comme vrais :

- $N^w p, p \rightarrow q$, donc $N^w q$. (O'Connor)
- $N^w p, N^w(p \rightarrow q)$, donc $N^w q$. (McKay et Johnson)

Deuxièmement, les versions de TC et de TI qui sont réfutées par des contre-exemples sont toujours plus ou moins éloignées de l'idée exprimée par l'argument du transfert. Par exemple, lorsqu'une version de TC utilise, dans sa deuxième proposition, une simple relation d'implication (donc « $p \rightarrow q$ »), on réfute cette version au moyen d'exemples dans lesquels q est antérieur à p , ou est une tautologie. Or il est évident que ces contre-exemples sont très éloignés de l'idée exprimée par l'argument du transfert. En effet, comme celui-ci vise le déterminisme causal, il évoque une relation causale allant de p à q , où p est donc antérieur à q , et où aucune de ces deux propositions n'est logiquement nécessaire. De même, lorsqu'on réfute une version de TC dans laquelle l'opérateur A exprime la capacité large, on est loin de l'idée exprimée par l'argument du transfert. Ainsi, la réfutation de ces différentes versions ne jette pas le doute sur le principe du transfert tant qu'il existe une ou deux versions de ce principe qui échappent aux contre-exemples et qui expriment de manière suffisamment simple et intuitive l'idée centrale de l'argument incompatibiliste. Tel est le cas du principe TI, facilement applicable au déterminisme causal. Le déterminisme causal est la thèse selon laquelle, en vertu des lois de la nature, si P , alors P cause F . Puisque les lois de la nature sont fixes (au moins au sens du principe FLC), le déterminisme implique qu'on peut faire le raisonnement suivant :

1. $N^c P$
2. En vertu des lois de la nature, si P , alors P cause F
3. P cause F (en vertu de (1) et (2))
4. $N^c P$ et (si P , P cause F), alors $N^c F$ (application de TCI)
5. $N^c F$ (en vertu de (1), (3) et (4)).

Mais comme nous l'avons vu plus haut³⁶, un compatibiliste peut admettre un tel raisonnement sans renoncer à sa position, car la conclusion ne porte pas directement sur la responsabilité morale. Puisque TCI n'est pas le transfert de non-responsabilité (TNR), l'argument que je viens d'exposer ne prouve pas directement l'incompatibilité du déterminisme causal avec l'idée que nous sommes responsables d'une partie de nos actions. Mais inversement, si TNR ne dépend pas étroitement de TCI, alors les objections contre TCI ne nous empêchent pas d'employer TNR dans une version de l'argument du transfert portant explicitement sur la notion de responsabilité. Il en découle que, pour résister à l'argument du transfert, la meilleure réponse compatibiliste et la plus directe est celle qui réfute TNR au moyen d'exemples. Fischer, Ravizza et Stump, ont proposé de tels exemples. Nous allons voir maintenant s'ils réfutent effectivement TNR.

3.7 Le transfert de non-responsabilité (TNR)

Le principe le plus pertinent pour établir la thèse incompatibiliste est le principe du transfert de non-responsabilité, car il porte directement sur la relation entre déterminisme et responsabilité morale. Comme TI et TC, il admet plusieurs formulations. Considérons d'abord ses formulations les plus simples. Posons que « Rp » signifie : « quelqu'un est responsable (au moins en partie) de p ». TNR peut d'abord se présenter ainsi :

(TNR1) Si $\sim Rp$, et si $p \rightarrow q$, alors $\sim Rq$.

(TNR2) Si $\sim Rp$, et si $\sim R(p \rightarrow q)$, alors $\sim Rq$.³⁷

L'argument se construit de la même manière qu'avec le connecteur « N ». Cependant, il faut modifier la première prémisse, car on peut être responsable d'un événement appartenant au passé actuel. Si p se produit à t1, alors, après t1, je ne peux plus empêcher « p se produit à t1 », mais je peux tout de même en être responsable. Par précaution, nous pouvons donc

³⁶ Cf. *supra*, p. 133.

³⁷ Ces principes correspondent aux versions du transfert d'incapacité proposées par O'Connor (pour la première), et par McKay et Johnson. Cf. les principes TCI 2 et TCI 3, dans la section précédente, p. 144.

redéfinir P comme l'ensemble des événements et des états de choses du passé actuel à *un moment antérieur à l'origine de l'humanité*. Si le déterminisme est vrai, $(P \wedge L) \rightarrow F$, donc on peut raisonner ainsi :

1. $\sim RP$
2. $\sim RL$
3. $\sim R(P \wedge L)$ en vertu de (1) et (2)
4. $\sim R((P \wedge L) \rightarrow F)$. Donc :
5. $\sim RF$ en vertu de (3) et de (4).

Les prémisses (1) et (2) paraissent difficilement contestables. Les arguments utilisés contre les principes de fixité FL et FP ne sont pas applicables à la responsabilité morale, car même si je suis capable de faire une action telle qu'une loi de la nature est violée à un moment antérieur, ou telle qu'un événement appartenant au passé actuel ne s'est pas produit, je ne suis certainement pas responsable des lois de la nature ni des événements ou des états de choses antérieurs à ma naissance. Il en découle apparemment que si le déterminisme est vrai, je ne suis jamais responsable de mes actions. Pour réfuter cet argument, Fischer, Stump, et Ravizza, ne critiquent pas les prémisses de l'argument, mais le principe du transfert de non-responsabilité, qui permet de passer de (3) et (4) à (5). Mark Ravizza³⁸ imagine la situation suivante :

[Betty] À t1, Betty fait exploser une charge de dynamite afin de provoquer une avalanche et d'écraser le camp ennemi sous la neige. Cette explosion détruit effectivement le camp ennemi. Or, à l'insu de Betty, sur les hauteurs enneigées, une brebis a heurté une pierre qui, en roulant, aurait de toute façon provoqué l'avalanche et entraîné la destruction du camp si Betty n'avait pas enfoncé la manette du détonateur. On suppose donc que cet événement (« la brebis heurte la pierre ») aurait été une cause suffisante de l'avalanche si Betty n'avait pas agi. Or personne n'est responsable du fait

³⁸ Mark Ravizza, "Semi-Compatibilism and the Transfert of Nonresponsibility", *Philosophical Studies* 75 (1994). Cet exemple est réutilisé par J. M. Fischer et E. Stump dans "Transfer Principles and Moral Responsibility". *Philosophical Perspectives* 14 (2000), p. 47-56.

que la brebis heurte la pierre à t_1 , et personne n'est responsable du fait que cet événement aurait entraîné l'avalanche puis la destruction du camp. Néanmoins, Betty est responsable de la destruction puisqu'elle l'a causée en provoquant l'explosion. TNR serait donc faux.

Remarquons d'abord que dans cet exemple, il n'y a pas de causalité simultanée. Pour cela, il faudrait que deux causes indépendantes provoquent le même événement, et que chacune, en l'absence de l'autre, soit suffisante pour le provoquer. Or l'événement « la brebis heurte la pierre à t_1 » est causalement suffisant pour provoquer l'avalanche, mais ne la cause pas. Il faut distinguer ici les relations d'*implications* et les relations *de cause à effet*. Le fait que la brebis heurte la pierre à t_1 *implique* la destruction du camp, mais ne la *cause* pas. Betty est donc pleinement responsable de la destruction du camp, même si elle ne peut l'empêcher.

Le principe TNR serait donc faux. Il n'aurait qu'une apparence de vérité due à sa ressemblance avec le théorème « si $\Box p$ et si $\Box(p \rightarrow q)$, alors $\Box q$ » (principe du transfert de nécessité). L'exemple de Betty semble apporter la preuve qu'on peut être responsable d'un fait (ici, l'avalanche) sans avoir la possibilité (ni la capacité) de l'empêcher : Betty ne peut empêcher l'avalanche, puisque si elle n'appuyait pas sur le détonateur, la pierre mue par la brebis suffirait à la provoquer.

Face à un tel contre-exemple, il y a au moins deux manières de défendre TNR. La première consiste à remplacer l'implication par une relation de causalité dans la deuxième proposition du transfert. Dans la section 3.6, j'ai appliqué un tel procédé au transfert d'impuissance (TI) pour obtenir un transfert *causal* d'impuissance qui évite les contre-exemples et peut être dérivé d'une version correspondante de TC (le transfert de capacité). En faisant la même substitution dans le cas du principe TNR, on le renforce également face aux objections compatibilistes. Pour cela, il faut d'abord revenir sur la définition du déterminisme. Dans les arguments incompatibilistes, on définit souvent le déterminisme de la manière suivante :

(déf.) Le déterminisme est la thèse selon laquelle P et L impliquent F (l'ensemble des événements passés, et l'ensemble des lois de la nature, impliquent l'ensemble des événements présents et futurs).

Nous pouvons envisager une autre définition, où le caractère causal du déterminisme est mis en évidence :

(déf.) Le déterminisme est la thèse selon laquelle tout événement présent ou futur est causé de manière nécessaire par des événements passés en vertu des lois de la nature. Autrement dit, P cause nécessairement F en vertu de L (l'ensemble des événements passés cause l'ensemble des événements présents ou futurs en vertu des lois de la nature).

De même, TNR peut se reformuler ainsi :

(TCNR) $\sim R_p$ et $\sim R(Cpq)$ impliquent $\sim Rq$. (Si personne n'est responsable de p, et si personne n'est responsable du fait que p cause q, alors personne n'est responsable de q).

Voici maintenant comment on peut répondre à l'objection de Ravizza, Fischer et Stump, à l'aide de TCNR ou d'une autre version causale de TNR. Beaucoup d'exemples utilisés dans la discussion du principe des possibilités alternatives mentionnent un intervenant contrefactuel, c'est-à-dire une cause suffisante pour provoquer l'événement, mais qui n'agit pas dans le monde actuel. Dans l'exemple de Betty, la brebis agit dans le monde actuel, mais ce n'est pas son action qui cause l'événement. Lorsqu'il emploie ce genre d'exemples pour attaquer le principe des possibilités alternatives, Fischer insiste sur le fait que puisque cette cause n'intervient pas (ou n'est pas la cause) dans le monde actuel, sa présence ou son absence ne change rien à la manière dont l'événement est produit, ni à la responsabilité de l'agent à l'égard de cet événement. Par exemple, si Gunnar tue Ridley sans que le savant fou ait besoin d'intervenir pour provoquer ce meurtre, Gunnar est bien responsable de la mort de Ridley. Or dans la discussion du transfert de non-responsabilité, le fait que la cause préemptive (qui rend l'événement inévitable) n'intervienne pas (ou ne cause pas l'événement) dans le monde actuel paraît se retourner cette fois contre la position de Fischer. Dans l'exemple de l'avalanche, l'événement « la brebis heurte la pierre à t1 » ne provoque pas l'avalanche dans le monde actuel. Betty la provoque elle-même, et c'est pourquoi elle est responsable de la destruction du camp. Mais en admettant que cet exemple réfute les principes TNR1 et 2, il ne peut pas réfuter TCNR. Pour cela, il ne suffit pas que l'événement « la brebis heurte la pierre à t1 » *implique* la destruction, mais il faut que cet événement la *cause*. Ainsi, pour que l'exemple de

l'avalanche réfute TCNR, il faudrait qu'une cause échappant au contrôle de Betty rende la destruction inévitable *en intervenant dans le monde actuel*.

Si l'avalanche était provoquée par la brebis, nous pourrions tenir le raisonnement suivant en appliquant TCNR:

- (1) $\sim R$ (la brebis heurte la pierre à t1).
- (2) $\sim R$ (l'événement « la brebis heurte la pierre à t1 » *cause* l'événement « le camp ennemi est détruit par une avalanche à t2 »). Donc :
- (3) $\sim R$ (le camp ennemi est détruit par une avalanche à t2).

Mais puisque la cause de l'avalanche est l'action de Betty, la proposition (2) est fausse, ce qui explique pourquoi Betty peut être responsable³⁹. TCNR demeure donc intact. Il est immunisé contre tous les exemples à la Frankfurt, car il est impossible de le critiquer en imaginant des situations contrefactuelles où une autre cause que Betty provoque l'avalanche. La formulation précédente autorisait ce genre de critique parce que la relation exprimée en (2) était une relation d'implication. A partir du moment où on la remplace par une relation de causalité, on ne peut plus parler que du monde actuel, de ce qui s'est réellement déroulé.⁴⁰

3.8 Les cas de causalité simultanée

Or Fischer a trouvé une réponse subtile à cet argument. Il s'appuie sur des exemples de causalité simultanée⁴¹. « Betty » est un cas de « surdétermination par préemption » (« *preemptive overdetermination* ») : il contient deux causes suffisantes pour produire la destruction du camp, mais une seule provoque cet événement, la deuxième étant seulement invoquée pour montrer que l'événement était inévitable. Or selon Fischer, si TNR (ou TCNR)

³⁹ J'admets en effet que $\sim R_p$ implique R_p , comme N_p implique N_p (selon la notation de Van Inwagen).

⁴⁰ TCNR est également immunisé contre tous les exemples où q désigne un événement antérieur à p , comme dans l'argumentation de David Lewis. En effet, dans une relation de causalité, l'effet est nécessairement postérieur à la cause, alors qu'une relation d'implication peut aller aussi bien du futur vers le passé que du passé vers le futur.

⁴¹ J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 159-169.

peut résister à ce genre de contre-exemples, il est manifestement faux dans les cas de « surdétermination simultanée » (« *simultaneous overdetermination* »), où l'événement est produit par deux causes à la fois.

Par exemple, dans « Assassins simultanés »⁴², Sam et Jack détestent le maire depuis qu'il impose une augmentation systématique des taxes. Un jour où se présente une occasion favorable pour éliminer le maire, Sam et Jack tirent sur lui au même moment. Il perd du sang rapidement et meurt. Or si Sam n'avait pas tiré, il serait mort également, car chacun des deux coups de feu était une cause suffisante pour le faire mourir. Le maire aurait perdu son sang moins vite, mais les secours seraient arrivés trop tard pour le sauver. Ici,

- Sam n'est pas responsable du fait que Jack tire.
- Sam n'est pas responsable du fait que l'événement « Jack tire sur le maire à t1 » provoque la mort du maire à t2.
- Pourtant, Sam est responsable de la mort du maire, car il l'a causée lui aussi.

De la même façon, en reprenant « Betty », on peut imaginer deux avalanches simultanées qui dévalent deux versants opposés et écrasent le camp ennemi au même moment. L'une des deux est provoquée par Betty, et si Betty n'avait pas appuyé sur le détonateur, l'autre avalanche aurait suffi pour le détruire. Betty n'est pas responsable de l'autre avalanche, ni du fait que cette avalanche causait inévitablement la destruction du camp ennemi. Pourtant, elle est responsable de la destruction du camp.

L'intérêt de ces exemples, du point de vue de Fischer, est qu'il est impossible d'utiliser contre eux l'argument qui a servi plus haut contre l'exemple de Ravizza (« Betty »). Dans le cas des assassins simultanés, et dans celui des avalanches simultanées, l'agent est responsable de l'événement bien qu'il soit *causé dans le monde actuel* par une cause échappant à son contrôle et agissant de façon inéluctable. TCNR paraît donc faux. Selon Fischer, la plausibilité de TNR et de ses versions améliorées vient du fait qu'il s'applique dans la plupart

⁴² J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 117-118.

des cas, mais il est possible de montrer qu'il ne vaut pas toujours. Si l'argumentation de Van Inwagen en faveur de TNR paraît convaincante, c'est parce qu'il utilise des exemples où TNR est manifestement valable, par exemple :

1. Personne n'est responsable du fait qu'un séisme d'un certain type se produit dans l'océan pacifique.
2. Personne n'est responsable du fait que si un tel séisme se produit dans l'océan pacifique, il y a ensuite un raz de marée sur la côte californienne.
3. Personne n'est responsable du fait qu'il y a un raz de marée sur la côte californienne.⁴³

Ici, personne ne peut causer le raz de marée en même temps que le séisme, c'est pourquoi personne ne peut en être responsable, mais dans un exemple où l'événement est plus facile à produire (la mort d'une personne, ou une avalanche), il est plus facile de trouver un contre-exemple à TNR (ou à TCNR).

Il y a pourtant un moyen de répondre à l'objection de Fischer. En effet, dans les cas de surdétermination simultanée, les deux causes ne peuvent pas produire *exactement* le même effet. Par exemple, dans « avalanches simultanées », s'il est vrai que les deux avalanches produisent en même temps la destruction du camp, elles ne produisent cependant pas la même partie de cet événement. Imaginons que l'avalanche produite par Betty vienne par le nord sur le camp ennemi, alors que l'autre avalanche dévale le versant opposé. Nous pouvons dire que Betty détruit seulement la partie nord du camp. Elle n'est pas responsable de toute la destruction. Puisque sa responsabilité n'est que partielle, l'exemple des avalanches simultanées n'est pas décisif contre TCNR. Il est encore possible de dire ceci :

- Betty n'est pas responsable de l'avalanche qui commence à t1 au sud.
- Betty n'est pas responsable du fait que si une avalanche commence à t1 au sud, la partie sud du camp est détruite à t2.
- Betty n'est pas responsable du fait que la partie sud du camp est détruite à t2.⁴⁴

⁴³ *Ibid.*, p. 164.

⁴⁴ *Ibid.*, p. 166.

Fischer utilise un exemple où deux causes provoquent la mort d'un individu nommé Jean le jour de ses trente ans : un cobra, par une morsure vénéneuse, et un assassin, par un coup de pistolet. Là encore, on imagine que chacune des deux causes est suffisante pour provoquer la mort de Jean. L'assassin est responsable de la mort de Jean même si la morsure vénéneuse du cobra la causait *de toute façon*. Mais nous pouvons répondre que les deux causes provoquent des événements différents : le coup de pistolet provoque une perte de sang importante, et la morsure provoque l'empoisonnement du sang qui reste dans le corps de Jean. L'assassin a évidemment une responsabilité dans la mort de Jean, et nous pouvons penser qu'il est blâmable au même degré que s'il avait provoqué tout seul la mort de Jean, mais il n'est pas entièrement responsable de ce qui s'est passé. Par exemple, la mort est survenue beaucoup plus vite que si le cobra n'avait pas mordu Jean ; c'est la conjonction des deux causes qui a provoqué une mort aussi rapide.

Une même interprétation s'applique enfin à l'exemple des assassins simultanés, bien qu'elle soit moins évidente dans ce cas : Sam et Jack provoquent bien tous les deux la mort du maire, mais en provoquant des événements différents. Par exemple, la balle tirée par Sam a touché la jambe gauche du maire, alors que celle tirée par Jack a atteint l'épaule droite de la victime. Sam est *partiellement* responsable de la mort du maire, et *entièrement* responsable de l'hémorragie provoquée par son coup de feu dans la jambe gauche du maire. Les deux événements qui provoquent la mort du maire sont du même type (ce sont des hémorragies), mais ils ne sont pas identiques pour autant.

Pour bien voir que cette réponse à l'objection de Fischer n'est pas purement *ad hoc*, il faut imaginer des situations où l'agent a une responsabilité moins grande, parce qu'il ne cause qu'une petite partie de l'événement. Par exemple, imaginons que l'avalanche provoquée par Betty détruise seulement un dixième du camp ennemi, tout le reste étant détruit par une autre avalanche provoquée par une brebis (ou une cause naturelle comme l'érosion d'un glacier). Imaginons de plus que la partie du camp détruite par Betty soit peu importante du point de vue stratégique, et qu'il n'y ait personne en cet endroit au moment où la neige écrase le camp. Dans ce cas, Betty est aussi blâmable que si elle avait détruit le camp à elle seule, mais sa responsabilité dans la destruction du camp ennemi est très limitée. Elle est néanmoins responsable d'une partie de l'événement. Ici, nous voyons qu'il est parfaitement légitime de

décomposer un événement en plusieurs parties pour attribuer à chacune de ses causes la responsabilité d'une de ses parties. Cela justifie une autre version de TNR :

(TCNR*) $\sim R_p$ et $\sim R(p \text{ cause entièrement } q)$ impliquent $\sim Rq$.

La seule différence entre TCNR* et TCNR est le mot « entièrement » ajouté dans la deuxième proposition du transfert. TCNR* s'applique au déterminisme en vertu de la définition suivante :

(déf.) Le déterminisme est la thèse selon laquelle pour tout événement y , il y a un événement a ou un ensemble d'événements $a-b-c$ tel que a , b , et c causent *entièrement* y de façon nécessaire. Autrement dit, P cause *entièrement* F en vertu de L .

Compte tenu de ces précisions, je pense que le principe du transfert de non-responsabilité est parfaitement valable. En effet, si le déterminisme est vrai, tous les événements, y compris nos actions libres, sont causés *entièrement* par des événements antérieurs à notre naissance. Cela paraît exclure la notion de responsabilité morale. Si aucun exemple ne réfute TCNR*, il est permis de croire que sa plausibilité ne repose pas sur une classe particulière d'exemples, mais qu'il est valable dans tous les cas possibles.

En fait, il n'est même pas nécessaire de recourir à TCNR*, car nous pouvons défendre TNR (et TCNR) en soutenant que les événements sont des particuliers et non des universels. Nous avons vu comment Van Inwagen défend le principe des possibilités alternatives contre les exemples à la Frankfurt : il montre que lorsqu'un individu provoque un événement en accomplissant une action, cet événement doit être envisagé comme un particulier⁴⁵. Par exemple, si Gunnar tue Ridley sous le contrôle d'un intervenant contrefactuel, il le tue par lui-même sans avoir la possibilité d'agir autrement. Dans ce cas, il est responsable de la mort de Ridley parce que même s'il ne pouvait pas éviter de tuer Ridley, il avait la possibilité d'éviter l'événement *particulier* qui se produit et que nous pouvons décrire ainsi : « Gunnar tue Ridley de lui-même ». Cet événement se distingue de l'événement « Gunnar tue Ridley sous l'influence du savant fou ». Gunnar n'est pas responsable de l'événement « Ridley meurt »

⁴⁵ Cf. *supra*, p. 66 et s.

compris comme un universel, car tout homme finit nécessairement par mourir un jour. Il ne peut donc être responsable que d'un événement particulier, caractérisé par une certaine *histoire causale*. C'est pourquoi « Ridley meurt » n'est pas le même événement dans la situation actuelle, où Gunnar la provoque sans y être déterminé causalement, et dans l'autre situation, où le savant fou provoque artificiellement en Gunnar le choix de tuer Ridley. De même l'action de Gunnar dans la situation actuelle n'est pas la même que si elle était déterminée causalement par le savant fou. Ainsi, nous pouvons dire que Gunnar a la possibilité d'agir autrement malgré la présence de l'intervenant contrefactuel. C'est pourquoi il est bien responsable de la mort de Ridley.

Or cette réponse de Van Inwagen s'applique de nouveau ici, dans la défense du principe TNR : Betty est bien responsable de l'avalanche, mais contrairement à ce que pensent Ravizza, Fischer, et Stump, elle avait la capacité d'éviter l'avalanche. Ou pour mieux dire, elle avait la possibilité d'éviter *cette avalanche* qu'elle a causée, même si elle n'avait pas la possibilité d'empêcher qu'il y eût une avalanche. Car si elle n'avait pas appuyé sur le détonateur, *cette avalanche* ne se serait pas produite ; mais il y aurait eu une *autre* avalanche, causée par la brebis. Cette avalanche est bien un autre événement : elle n'a pas les mêmes causes, et ne commence pas au même endroit. Sa ressemblance avec la première ne doit pas nous égarer ; elle tient au fait que les deux avalanches commencent à peu près au même moment et produisent des effets identiques à certains égards (surtout en ce qu'elles détruisent le camp ennemi). Mais en examinant la chose de plus près, on voit bien qu'il s'agit de deux événements différents. L'argument de la conséquence est donc valable.

Même dans l'exemple des deux assassins, les événements ne sont pas les mêmes. Dans le monde actuel, Sam et Jack tuent la victime simultanément. L'autre possibilité est qu'ils ne tuent pas la victime mais qu'elle meure tout de même, assassinée par Jack. Dans les deux cas, la victime meurt assassinée au moyen d'une arme à feu, et le moment de sa mort est à peu près le même dans les deux cas. Or en imaginant que ce moment soit exactement le même, les deux événements se distinguent encore par leur origine causale. L'argument de Fischer ne vaut donc que si le critère d'identité entre les deux événements est large, c'est-à-dire s'il faut seulement que les deux événements tombent sous la même description générale (« un tel meurt à t_2 », ou « un tel meurt assassiné à t_2 », ou « un tel meurt assassiné par balle à t_2 »,

etc.). Si ce critère était suffisant, nous pourrions imaginer des événements nettement différents sans modifier notre jugement sur la responsabilité de l'agent. Or ce n'est pas le cas. En augmentant la différence entre les deux événements, on voit que même s'ils tombent sous la même description générale, nous ne pouvons pas les considérer comme identiques. Par exemple, imaginons que Sam tue la victime par balle mais que Jack l'ait déjà empoisonnée avec des champignons en lui servant des bouchées à la reine. Comme les champignons agissent lentement sur l'organisme, la victime n'est pas encore morte au moment où Sam décide de la tuer d'un coup de revolver. Si Fischer a raison, un tel cas ne diffère pas vraiment du précédent, même si dans le cas présent, la différence entre les deux événements possibles est plus grande (soit la victime meurt par balle, soit elle meurt empoisonnée). En augmentant cette différence, on voit que la mort de la victime dans les deux scénarios n'est pas le même événement, mais deux événements différents. En poussant l'idée jusqu'au bout, on aboutit à des événements aussi différents que « Ridley est assassiné à 30 ans » et « Ridley meurt d'un arrêt cardiaque à 90 ans ». L'argument du transfert est donc valable. Il l'est parce que l'événement dont il s'agit dans TNR (ou dans TCNR) doit être compris comme un événement *singulier*, bien qu'il soit nommé à l'aide d'une description générale.

Résumé. Dans ce chapitre, j'ai discuté trois réponses différentes à l'argument du transfert. La première, que j'ai examinée dans les sections 3.2 et 3.3, consiste à nier une des prémisses de l'argument, c'est-à-dire un des principes de fixité : FL (fixité des lois de la nature) ou FP (fixité du passé). Nous avons vu que ces prémisses sont solides, particulièrement dans leurs versions causales. Or ce sont les versions causales qui importent dans l'argument du transfert, car la capacité de faire que non-p (ou d'empêcher que p) implique un contrôle causal et non seulement la capacité de faire une action telle que si on l'accomplit, p est fausse.

Dans la section 3.4, j'ai abordé la deuxième réponse compatibiliste à l'argument du transfert. Elle consiste à nier sa validité en montrant que les règles de clôture valant pour la nécessité logique ne valent pas pour la notion d'incapacité (ou la « *power-necessity* »). Mais cette réponse de Michael Slote n'est pas assez convaincante, car elle suppose que la « *power-*

necessity » ressemble à la non-accidentalité, ce qui n'est pas le cas. La non-accidentalité d'un événement, telle que la comprend Slote dans ses contre-exemples, dépend des désirs et des pensées de certains individus. En cela, elle se rapproche bien plus de la nécessité épistémique que de l'impuissance ou « *power-necessity* », notion objective exprimant une nécessité aléthique.

Comme nous l'avons vu en 3.5, la seule possibilité qui reste au compatibiliste pour résister à l'argument du transfert est de soutenir que même s'il est valide, il ne prouve pas l'incompatibilité du déterminisme avec la responsabilité morale, mais seulement l'incompatibilité du déterminisme avec un certain pouvoir des alternatives. Mais pour que cette idée ne soit pas une simple hypothèse, le compatibiliste doit prouver que l'argument du transfert ne fonctionne pas lorsqu'il porte directement sur la non-responsabilité (c'est-à-dire lorsqu'on remplace le connecteur « N » par l'opérateur de non-responsabilité « $\sim R$ »). Pour cela, le compatibiliste doit réfuter TNR, le principe du transfert de non-responsabilité. Or nous venons de voir que les arguments utilisés contre TNR ne le réfutent pas.

L'argument du transfert est donc a priori valide et concluant lorsqu'il utilise le l'opérateur de non-responsabilité pour prouver directement la thèse incompatibiliste. Or, avant d'aborder TNR, nous avons vu que l'argument du transfert est valide avec le connecteur d'impuissance. Nous pouvons donc penser que si le schéma de raisonnement appelé « argument du transfert » est valide avec le connecteur de non-responsabilité, c'est parce qu'il l'est aussi avec le connecteur d'impuissance. Enfin, nous avons vu dans la section 3.6 que le principe du transfert d'impuissance (TI), sous sa forme causale (TCI), pouvait être dérivée de TCC : le transfert causal de capacité. Nous disposons ainsi de trois principes de transfert *causaux* qui semblent étroitement liés : le transfert causal de capacité (TCC), le transfert causal d'impuissance (TCI), et le transfert causal de non-responsabilité (TCNR). Le fait qu'ils forment ensemble un système cohérent et qu'ils tiennent tous les trois face aux exemples les plus variés suggère qu'ils sont vrais. Je pense donc que ces principes sont corrects et constituent un fondement solide de la thèse incompatibiliste.

Dans les prochaines sections, j'essaierai d'énoncer les principales conditions de la responsabilité morale, pour montrer ensuite qu'elles peuvent se ramener à une seule condition

: un agent *A* est responsable d'un événement *E* seulement s'il est une source ultime de *E*, c'est-à-dire seulement si l'action (ou l'omission) par laquelle il cause *E* n'est pas elle-même causée. Ainsi unifiée, la conception libertarienne de la responsabilité morale se caractérise par une grande cohérence. Dans les chapitres IV à VI, je soutiendrai que les théories compatibilistes, au contraire, ne parviennent pas à tenir compte de manière cohérente et intelligible de toutes nos intuitions sur les conditions de la responsabilité morale.

3.9 Première condition de la responsabilité : la contribution de l'agent

Nous avons vu dans les sections précédentes que les deux principes sur lesquels Van Inwagen a fondé sa critique du compatibilisme (le principe des possibilités alternatives, et le principe du transfert d'impuissance) résistent finalement aux objections qu'on leur oppose de toutes parts. Le premier, cependant, ne résiste à la critique qu'en s'éloignant des intuitions qu'il paraissait exprimer au début. De prime abord, ce que nous appelons la capacité d'agir autrement, c'est la capacité d'agir de manière à éviter un certain état de choses qui s'est réalisé, en produisant un résultat différent. Par exemple, nous pensons que si un assassin était incapable de ne pas tuer sa victime, il n'est pas responsable de ce qu'il a fait. « Tuer la victime » est ici une description générale, car si l'assassin avait seulement la capacité de la tuer une minute plus tard, il ne paraît pas plus responsable. Or les exemples à la Frankfurt excluent que la responsabilité de l'agent soit toujours liée à la possibilité d'amener un résultat différent. Finalement, la possibilité d'agir autrement se réduit à la possibilité de ne pas être la véritable cause de l'événement. Cette condition revient à dire que l'action ne doit pas être causée, sinon l'agent n'est pas responsable de l'événement. Le principe TCNR confirme cette idée : si l'action est l'effet nécessaire d'une cause non-contrôlée par l'agent, il n'en est pas responsable. Il ne suffit donc pas que l'agent soit une cause de l'événement, il faut également qu'il en soit une cause première, donc *autonome* au sens le plus fort. Les compatibilistes admettent volontiers que la liberté suppose l'autonomie, mais ils ne parlent pas d'une autonomie aussi forte. Avant d'examiner leurs théories, je vais d'abord essayer de montrer ce qui relie les différentes conditions de la responsabilité. Je soutiendrai qu'une conception forte

de l'autonomie de l'agent permet de rendre compte des autres conditions, et d'acquiescer ainsi une vision cohérente et unifiée de la responsabilité morale.

Notons pour commencer qu'on n'a pas besoin d'être une cause *nécessaire* d'un événement pour en être responsable. C'est ce que montrent les exemples utilisés contre TNR. Dans « Betty », Betty n'est pas une cause nécessaire de la destruction du camp ennemi, du moins si nous comprenons l'événement « la destruction du camp » comme un universel. Faut-il en revanche qu'elle soit une cause *suffisante* de l'événement ? Non plus. Imaginons que Betty ait besoin du concours de plusieurs personnes pour détruire le camp ennemi : un spécialiste des explosifs et un petit groupe de soldats exercés à l'escalade, pour installer la dynamite au bon endroit. Bien entendu, le fait qu'elle n'agisse pas seule ne la décharge pas de toute responsabilité. Prenons un autre exemple. Des soldats se demandent s'ils vont laisser en vie un prisonnier, ou s'ils vont s'en débarrasser en l'abandonnant dans un endroit désertique où il a peu de chances de survivre. Ils tirent à pile ou face ; si la pièce tombe sur pile, le prisonnier reste avec eux, si elle tombe sur face, ils l'abandonnent en ce lieu hostile. Or la pièce tombe sur face, et le prisonnier meurt de soif quelques jours après. Les soldats ne sont pas la *cause nécessaire* de sa mort, puisqu'il y a plusieurs manières différentes de mourir et que cet individu, de toute façon, serait mort un jour d'une manière ou d'une autre. Sont-ils la *cause suffisante* de sa mort ? Non plus, car le fait que la pièce tombe sur face ne dépend pas seulement d'eux, mais de facteurs physiques qu'ils ne contrôlent pas. Ensuite, la mort du prisonnier n'est pas directement causée par eux. Par exemple, il pourrait continuer à vivre s'il trouvait un puits ou s'il rencontrait une caravane. Les soldats ne sont donc pas une cause suffisante de sa mort, mais ils en sont responsables car ils y ont contribué, et ils pouvaient ne pas y contribuer.

On voit que la responsabilité d'un agent ne suppose pas qu'il soit une cause nécessaire, ni une cause suffisante du fait en question. En fait, il n'est même pas nécessaire d'agir, ni de *produire* quelque chose, pour être responsable d'un fait. Il arrive qu'on soit responsable d'un fait en omettant d'agir. On peut d'ailleurs être inculpé pour non-assistance à une personne en danger. Si en me promenant sur la plage, je vois un enfant qui se noie, je dois l'aider dans la mesure du possible. Si je reste passif alors que personne d'autre ne peut sauver l'enfant, je suis moralement responsable de la noyade. Pourtant, je n'ai rien fait, et ce n'est pas moi qui

l'ai noyé. Il est donc important de préciser que la contribution de l'agent à un événement dont il est responsable ne consiste pas toujours dans une action proprement dite ; elle peut également consister en une omission.

Pourtant, en dehors des cas d'omissions, le fait que l'agent cause l'événement semble être une condition très importante. Pour le voir, il faut imaginer des situations où l'agent fait quelque chose qui pourrait causer l'événement, mais sans parvenir à le causer. Par exemple, imaginons que Betty appuie sur le détonateur après le début d'une avalanche provoquée par une brebis. Cette fois, elle n'est pas responsable, alors qu'elle a tout fait pour provoquer l'avalanche. Elle est seulement blâmable ou méritante selon le point de vue auquel on se place. Pour bien voir la différence entre ces deux notions, comparons les deux situations suivantes. Un automobiliste qui percute un piéton involontairement est responsable de sa mort, sans être nécessairement coupable d'une faute grave. Par exemple, il peut arriver ceci : le piéton se jette sous ses roues, et dans l'obscurité de la nuit, l'automobiliste ne voit pas assez bien pour réagir à temps, mais on lui reproche d'avoir dépassé de 20 km/h les limitations de vitesse. Dans ce cas, l'automobiliste n'avait pas la possibilité de freiner à temps. Cependant, on suppose qu'il l'aurait eue s'il avait respecté les limitations de vitesse, et qu'il avait la possibilité de respecter les limitations de vitesse. Sa responsabilité est donc dérivée. On le considère comme responsable de la mort du piéton mais on ne le blâme pas pour l'avoir tué ; on le blâme d'avoir roulé nettement au-dessus des limitations de vitesse. A l'inverse, un automobiliste peut foncer sur un piéton avec une intention criminelle, mais manquer sa cible. Dans ce cas, il est blâmable au plus haut point, mais n'a rien causé. Il n'est donc pas responsable de la mort du piéton, puisque celui-ci n'est pas mort. Il est seulement responsable de sa mauvaise intention et d'une tentative d'assassinat.

Maintenant, imaginons le cas suivant : l'automobiliste fonce sur sa cible, mais la manque. Le piéton qu'il voulait écraser continue son chemin tranquillement, sans se douter de rien, mais vingt mètres plus loin, une autre voiture l'écrase accidentellement. Là encore, l'automobiliste qui avait des intentions criminelles n'est pas responsable de l'événement. De même, si on imagine que Betty n'est pas la vraie cause de l'avalanche mais appuie sur le détonateur juste après que la brebis a provoqué l'avalanche, elle n'en est pas responsable. Pour mieux voir cela, il faut imaginer un intervalle de temps plus long entre le moment où Betty

appuie sur le détonateur et celui où la brebis heurte la pierre. Supposons donc qu'il y ait une demie journée entre les deux événements (Betty et ses camarades sont mal renseignés ; l'ennemi a été écrasé par une avalanche il y a six heures). Les charges de dynamites ont été posées avant, et dans le camp de Betty, personne n'est au courant de l'événement. Là où elle se trouve, elle n'aperçoit pas le camp ennemi, ou plutôt ce qu'il en reste. Elle appuie donc sur le détonateur, mais en vain. Betty est dans la situation d'un homme qui frappe de toutes ses forces sur un cadavre dans l'intention de tuer. Il est évident qu'elle n'est pas responsable de la destruction du camp ennemi. Alfred Hitchcock a imaginé un cas de ce genre dans *Frenzy* : à la fin du film, le faux coupable s'échappe de sa prison. Pour se venger, il se dirige vers le lit du vrai coupable, muni d'une barre de fer. Il frappe, mais le corps allongé sous les draps est un cadavre.

Le paradoxe est donc là : on peut être responsable d'un événement sans avoir rien fait, et on peut aussi ne pas en être responsable alors même qu'on a tout fait pour le provoquer. Le deuxième point s'explique aisément à partir du moment où l'on distingue la responsabilité et le fait d'être blâmable. Pour être responsable d'un fait, il ne suffit pas de *s'efforcer* à le produire ; il faut le *causer* ou *y contribuer* réellement. Mais alors, on peut se demander pourquoi une omission est parfois considérée comme une *contribution* de l'agent. Pourquoi penser qu'un individu *contribue* à un événement lorsqu'il omet seulement de s'y opposer ? Il y a plusieurs manières de répondre à cette question. L'une d'elles consiste à dire que l'individu avait la possibilité ou la capacité d'agir autrement, ce qui implique à première vue une certaine contingence, au sens le plus fort du terme.

Précisons cette idée. Il y a des situations où une personne reste passive face un événement se produisant sous ses yeux, mais où elle n'en est pas responsable alors qu'elle a la capacité de faire quelque chose pour l'empêcher de se produire. Par exemple, un individu se promène sur une plage, et aperçoit un enfant en train de se noyer. Au lieu de se jeter à l'eau et de nager vers l'enfant pour le sauver, il poursuit négligemment sa promenade. Son comportement est évidemment condamnable, mais imaginons qu'à son insu, il y ait un requin qui avance vers l'enfant et qui le tuerait infailliblement si le promeneur essayait de le sauver, l'animal étant suffisamment près pour atteindre l'enfant en premier. Intuitivement, nous estimons que le promeneur n'est pas responsable du malheur, car il ne pouvait rien faire

d'efficace. La responsabilité dans une omission suppose donc non seulement la capacité de faire *quelque chose* pour éviter l'événement, mais la capacité de faire *quelque chose de suffisant* pour éviter l'événement dont on le juge responsable.

Ainsi apparaît un point commun entre les actions et les omissions : *ce que fait actuellement l'agent, ou ce qu'il aurait pu faire (dans les cas d'omissions) doit être efficace. Une tentative ne suffit pas (ni la possibilité de tenter quelque chose)*. Je reviendrai longuement sur le cas des omissions dans le chapitre V, en montrant que les conditions de la responsabilité dans les actions et dans les omissions ne sont pas tout à fait les mêmes.⁴⁶ Mais dans les deux cas, elle suppose l'autonomie ultime de l'agent et l'existence d'une certaine possibilité alternative. Nous verrons que l'agent n'est pas responsable d'un événement s'il n'a pas la possibilité de faire quelque chose d'efficace contre cet événement.

Résumons les remarques précédentes. Pour être responsable d'un événement E, il n'est pas nécessaire :

- d'être une cause suffisante de E
- d'être une cause nécessaire de E
- d'être la seule cause de E
- ni de causer E par une action.

Mais pour que S soit responsable de l'événement E, il faut que S cause E par une action, ou contribue à E par une omission.

Et S *contribue* à E par omission seulement s'il y a au moins une action A telle que :

- S peut faire A.
- Si S fait A, A empêche E de se produire.

⁴⁶ Pour Fischer et Ravizza, au contraire, la responsabilité dans les omissions a les mêmes conditions que la responsabilité dans les actions. Cf. J. M Fischer & M. Ravizza, *Responsibility and Control*, p. 131-150.

Je laisse de côté, pour l'instant, la question de savoir si l'on peut *causer* quelque chose par omission.⁴⁷ Ce qui importe, ici, c'est le fait que la responsabilité dans une omission suppose une possibilité alternative. Cette possibilité alternative suggère que le comportement de l'agent n'est pas déterminé causalement lorsqu'il est responsable de ce qu'il fait. Cela suggère aussi que le principe des possibilités alternatives exprime bien une condition de la responsabilité morale dans les actions. En effet, nous pouvons supposer que la responsabilité dans les actions et la responsabilité dans les omissions ne sont pas deux notions complètement distinctes, mais une seule notion qui prend deux formes partiellement différentes. Donc si la responsabilité dans les omissions suppose une capacité des alternatives, cela confirme notre impression que la responsabilité dans les actions suppose cette même capacité. A un niveau plus profond, l'idée d'*autonomie ultime* de l'agent permet de rendre compte à la fois des cas d'omissions et des cas d'actions. En effet, si l'agent possède une autonomie ultime dans ses comportements, il en découle une capacité de se comporter autrement. Lorsqu'il cause un événement sans avoir la possibilité de l'empêcher (au sens où il ne peut pas empêcher qu'un événement du même genre se produise), il en est responsable parce qu'il agit par lui-même. Et lorsqu'il ne cause pas un événement mais a la possibilité d'agir efficacement contre cet événement, il en est responsable parce qu'ayant l'autonomie ultime - donc la capacité de prendre des initiatives, il pouvait intervenir. Or l'agent n'aurait pas la possibilité causale d'agir autrement si son action ou son omission résultait d'une cause qu'il ne contrôle pas. Pour qu'un individu ait le pouvoir d'agir autrement, il faut donc que son comportement ne soit pas causé. Dans tous les cas, la responsabilité semble supposer que l'agent cause l'événement sans être causé, ou *permette* l'événement sans être causé. Ainsi, pour toute action A, toute omission O, et tout agent S, si S fait A (ou O),

S a le pouvoir de ne pas faire A (ou de ne pas faire O) ↔ L'action A de S (ou l'omission O de S) n'est pas causée.

⁴⁷ C'est une question débattue. Carolina Sartorio soutient qu'on peut causer un événement par omission. Cf. Carolina Sartorio, "A New Asymmetry Between Actions and Omissions", *Noûs* 39/3 (2005), p. 460-482.

La responsabilité morale semble liée à ces deux conditions indissociables. La contribution de l'agent à l'événement dont il est responsable signifierait alors qu'il contribue à E en étant une cause ultime ou une origine ultime de E.

3.10 La contribution volontaire de l'agent. Conditions épistémiques de la responsabilité morale

Comme nous venons de le voir, nul ne peut être tenu pour responsable d'un fait s'il n'a pas contribué à sa réalisation, et l'omission ou la passivité est une manière d'y contribuer. On s'accordera aisément sur une autre condition de la responsabilité : le caractère *volontaire* de cette contribution. Mais ce critère peut être mal compris.

Tout d'abord, il faut noter qu'une action peut être volontaire sans être *consciente*. Lorsqu'un individu conduit une voiture, il accomplit certaines actions par habitude, et sans s'en apercevoir. Par exemple, pour aller au travail ou faire ses courses, il tourne là où il faut, sans avoir besoin d'y penser. Cette action est inconsciente, mais volontaire. Il fait ces choses, non parce qu'il les veut en elles-mêmes, mais parce qu'il veut faire des courses ou aller au travail. Imaginons que je prenne ma voiture pour aller travailler. Le mouvement de mon bras, lorsque je tourne le volant ou que je manipule le levier de vitesse, est volontaire au sens où il est une partie d'une action plus globale que j'ai voulue : celle d'aller au travail. Quand un organiste improvise, son action se compose d'une multitude de gestes, le plus souvent inconscients, résultant d'habitudes acquises grâce à un entraînement plus ou moins long et méthodique. Lorsqu'il débute dans l'art de l'improvisation, il doit penser aux accords qu'il va jouer. Alors, on peut dire que chaque accord est voulu *consciemment*. Mais l'entraînement lui permettra ensuite de jouer ses accords plus vite et sans y penser. L'action de jouer un accord demeurera pourtant une action volontaire, car l'organiste aura voulu improviser, et il l'aura fait à sa guise, dans le style qu'il aura choisi.

Le principe suivant est donc douteux :

(P1) Une action A de S est volontaire seulement si elle résulte du choix de S de faire A.

Cependant, on peut soutenir que les actions habituelles ou inconscientes que je viens d'évoquer sont volontaires au sens où elles résultent de choix qui impliquent ces actions sans porter sur elles. Par exemple, le fait que je tourne à tel endroit est impliqué par mon choix d'aller au supermarché. Dans ce cas, même si je n'ai pas *décidé* de tourner à cet endroit, mon action de tourner est une conséquence d'un de mes choix. On peut donc proposer le principe suivant :

(P2) Une action A de S est volontaire seulement si elle résulte d'un choix de S.

Or même cette affirmation est trop restrictive, car il y a des comportements volontaires qui ne sont pas choisis et qui ne découlent pas d'un choix antérieur de l'agent. Par exemple, on peut élever le ton volontairement dans une conversation pour manifester de la colère ou impressionner son interlocuteur, sans avoir décidé de le faire et sans que ce comportement découle d'un choix précédent. De même, en faisant une promenade, on peut passer volontairement par tel endroit parce qu'on le trouve attirant à ce moment-là, et sans *choisir* de le faire. Ce comportement n'est pas nécessairement l'effet d'un habitus. Il peut venir d'une vague curiosité ou d'un désir plus ou moins inconscient de changer d'itinéraire.

Je pense donc qu'en dehors de certaines conditions épistémiques, une action est volontaire si l'agent a la *capacité d'agir autrement*. Aussi ne distingue-t-on pas une action *volontaire* d'une action *libre*. Les mouvements de l'organiste ou de l'automobiliste sont volontaires en ce sens. Pour un incompatibiliste, cette *capacité* d'agir autrement va de pair avec la *possibilité* d'agir autrement, mais nous avons vu qu'il fallait distinguer ces deux notions. Néanmoins, d'après les définitions que nous avons posées dans la première partie⁴⁸, la capacité stricte suppose l'absence d'empêchement, donc la possibilité au sens causal, qui n'existe pas dans la thèse déterministe. Il faut que ni cause, ni obstacle, ni défaut, n'empêche l'agent d'accomplir son action.

De plus, la thèse P2 se heurte au fameux problème des chaînes causales déviantes. Une chaîne causale déviante est une série de causes et d'effets par laquelle une volonté de l'agent cause accidentellement l'effet voulu. Par exemple, Gunnar veut tuer Ridley mais ne réussit pas

⁴⁸ Cf. *supra*, p. 28 et s.

à le toucher parce que Ridley est trop loin. Furieux, il monte dans sa voiture, démarre brusquement et dépasse largement les limitations de vitesse. La route fait une boucle près de l'endroit où il a essayé de tuer Ridley. Là, il aperçoit un piéton et freine tant qu'il peut pour l'éviter. Trop tard ! Il percute le piéton et s'aperçoit que c'est Ridley... Il a tué Ridley, ce qui était conforme à sa volonté, mais il ne l'a pas tué *volontairement*. D'ailleurs, il peut être mécontent de la façon dont il l'a tué (imaginons qu'un individu ait relevé le numéro de sa plaque d'immatriculation). Un causaliste ou un volitionniste ne peut pas se contenter de dire que l'action est volontaire si elle a été causée par une volonté (ou une volition) de l'agent, il faut aussi qu'elle ait atteint son but *sous le contrôle* de l'agent. Toute la difficulté sera de préciser en quoi consiste un tel contrôle.⁴⁹

Ces deux raisons suggèrent qu'il faut remplacer P2 par :

(P3) Une action A que S fait à l'instant t est volontaire si S peut ne pas faire A ou s'il a eu auparavant la possibilité de se comporter de manière à ne pas faire A à t.

P3 est une version du principe des possibilités alternatives appliquée à l'action volontaire. Il semble donc qu'en dehors des conditions épistémiques, le caractère volontaire d'une action ou d'une omission dont l'auteur est responsable ne suppose rien de plus que le principe des possibilités alternatives. Ainsi, pour énoncer toutes les conditions de la responsabilité morale, il ne sera pas utile de préciser que l'action (ou l'omission) doit être volontaire. Il suffira d'énoncer le principe des possibilités alternatives et de le compléter par des conditions épistémiques.

Il est en effet évident que la responsabilité morale dépend de certaines connaissances. Par exemple, imaginons un automobiliste qui écrase un piéton en traversant un lieu-dit. Supposons qu'il ne puisse pas du tout prévoir le risque qu'une personne se trouve sur la route au moment où il arrive. Dans ce cas, l'accident paraît entièrement contingent, et à cause de son caractère imprévisible, on excuse l'automobiliste (du moins, s'il respectait les limitations

⁴⁹ La notion de contrôle sera traitée dans les chapitre IV à VII. Pour Fischer, comme nous l'avons déjà vu dans le chapitre II, il existe une forme de contrôle qui ne suppose pas la capacité d'agir autrement : c'est le contrôle de direction ("*guidance control*"). Selon lui, ce contrôle suffit pour la responsabilité morale. Cf. *supra*, p. 65.

de vitesse). C'est ce qui arrive si en roulant sur l'autoroute en pleine nuit, on percute un individu qui se précipite sur la voie au dernier moment et de manière à ne pas être vu assez tôt pour qu'on l'évite. Dans ce cas, le piéton imprudent (ou suicidaire) est considéré comme le seul responsable de sa propre mort. L'automobiliste n'en est *pas responsable* s'il ne pouvait pas anticiper un tel risque. Au contraire, l'automobiliste serait *en partie* responsable s'il avait dépassé les limitations de vitesse, car il est censé savoir qu'en les dépassant, il perd sa capacité de réagir à temps en cas de nécessité.

Ainsi, pour être responsable d'un événement, l'agent doit être capable de savoir que son action peut entraîner cet événement. Autrement dit,

S est responsable de E si S est capable de savoir que E est une conséquence probable de son action.

Cette condition doit être complétée par une autre :

*S est responsable de E si S est capable de savoir si E est bon ou mauvais.*⁵⁰

En effet, si je sais que mon action A entraîne la conséquence E, mais en ignorant que E doit être absolument évité, je ne suis pas responsable de E. Par exemple, imaginons un enfant cueillant des fleurs dans un jardin public très soigné, ou son action est interdite. Il n'est pas responsable s'il n'est pas conscient du caractère mauvais de cet acte.

Ces deux conditions épistémiques posent un même problème lorsque l'agent possède les bonnes croyances au sujet de E, mais de manière accidentelle. Considérons d'abord la première condition. Que se passe-t-il si l'agent croit à tort que E est probable, et que son action entraîne pourtant E de façon accidentelle ? Imaginons qu'il emprunte la voiture d'un ami et la gare en lisière de forêt. Puis il s'éloigne en pensant à la voiture avec une certaine inquiétude, parce qu'il redoute à tort le passage d'un éléphant qui pourrait s'asseoir dessus (on suppose qu'il a vécu longtemps en Afrique). Ensuite, alors qu'il se promène dans les bois,

⁵⁰ Et S est sans doute responsable de E s'il sait que E est *probablement* mauvais, ou s'il sait que E est *probablement* bon.

l'événement se produit malgré son extrême improbabilité. A quelque distance de là, un éléphant s'est échappé d'un cirque.

Ici, on se demande si l'agent a une responsabilité dans l'incident. La même difficulté se pose à propos de la deuxième condition épistémique. Par exemple, imaginons un chasseur qui croit à tort se trouver en un lieu où il est interdit de chasser. Il chasse pourtant, agissant donc contre sa conscience. Bien que le fait de chasser en ce lieu ne soit pas *en soi* une action condamnable, le chasseur est blâmable. Il est responsable d'une action qui n'est pas mauvaise objectivement, mais subjectivement, en contribuant à un fait qui lui semblait mauvais. Mais ce cas est problématique. Le comportement de l'agent peut certes être l'objet d'une évaluation morale - on peut dire, en effet, qu'il s'est mal comporté, puisqu'il a agi contre sa conscience - mais est-il pour autant responsable de ce qu'il fait ? C'est douteux. Si nous sommes tentés de lui attribuer une responsabilité morale, c'est sans doute parce qu'il est responsable de son intention, mais il n'est pas responsable son action.

Dans d'autres cas où l'agent se trompe sur la valeur de son action, l'indépendance de ces deux éléments apparaît plus clairement. Imaginons un enfant cueillant des fleurs pour sa mère dans les jardins du château de Versailles, à vingt mètres de la façade principale. L'intention est bonne, mais l'acte est mauvais. On peut donc louer l'enfant pour sa bonne intention, et dire en même temps que son acte est mauvais, mais on ne peut pas le considérer comme responsable moralement de la dégradation du parterre. Dans ces deux exemples, l'agent est moralement responsable de son intention, mais il n'est pas moralement responsable de son action, ni des conséquences de son action.

Résumons maintenant les conditions de la responsabilité morale à l'égard d'un événement. J'énoncerai ici les conditions de la responsabilité dans une action (RA), en supposant qu'elles sont à peu près valable dans les omissions. Au chapitre V, j'examinerai plus à fond le cas des omissions.

(RA) S est moralement responsable de l'événement E en faisant l'action A si et seulement si :

(a) S contribue à E par l'action A.

Il faut aussi que A soit une action volontaire. Nous avons vu que son caractère volontaire ne suppose pas qu'elle soit choisie, ou voulue *consciemment*. En revanche, elle suppose les trois conditions suivantes :

(b) S peut ne pas faire A.

(c) S peut savoir si E est bon ou mauvais.

(d) S peut savoir que E est une conséquence probable de A.

Il est intéressant d'observer que toutes ces conditions paraissent liées à l'autonomie de l'agent. Dans le cas particulier des omissions, la condition (a) signifie a priori que S pourrait faire une action empêchant E de se produire, mais n'accomplit pas cette action dans la situation actuelle. Donc dans le cas des omissions, (a) suppose comme (b) que l'agent ait une capacité des alternatives. Or cette capacité, pour un libertarien, suppose que le comportement de l'agent ne soit pas causé. J'ai déjà expliqué ce lien entre la contingence et l'absence de cause :

(NC) Si x et y sont des substances ou des événements, alors (x rend nécessaire $y \rightarrow x$ cause y).⁵¹

On en déduit, par contraposition :

Si x et y sont des substances ou des événements, alors (y n'est pas causé par $x \rightarrow y$ n'est pas rendu nécessaire par x).

⁵¹ Cf. *supra*, p. 44. Il faut comprendre « x » et « y » comme des particuliers, et non comme des universels. Sinon, ce principe est réfuté par les exemples à la Frankfurt. Rappelons que dans ces exemples, la présence de l'intervenant contrefactuel rend nécessaire l'action (comprise comme un universel), mais sans la causer.

Donc l'autonomie ultime que possède un agent en faisant *y* suppose normalement que *y* ne soit rendu nécessaire par aucun *x* (substance, événement ou état de chose). Ainsi l'autonomie implique une certaine contingence à la source de l'acte, cette contingence qu'exprime le principe des possibilités alternatives.

Or l'autonomie ultime implique également les conditions épistémiques de la responsabilité morale. Elle suppose que l'agent accomplisse son action de façon consciente. Il n'est pas nécessaire qu'il agisse en ayant *actuellement* conscience de ses moindres mouvements, mais il faut au moins qu'il sache *en puissance* ce qu'il fait (c'est-à-dire *puisse s'apercevoir* de ce qu'il fait). En effet, en dehors de ces conditions, les mouvements de l'agent ne sont pas des actions qu'il accomplit par lui-même. Les compatibilistes eux-mêmes admettent généralement que l'autonomie requise par la responsabilité morale est liée à la rationalité de l'agent. Lorsqu'il agit dans l'ignorance, ses choix ne dépendent pas de motifs rationnels, mais d'influences extérieures. Ces influences sont très diverses, car elles peuvent être aussi bien des stimulations sensorielles que des influences sociales ou culturelles. Si l'agent n'a pas les connaissances requises par les conditions (c) et (d), il perd son autonomie en subissant ces influences. Si l'autonomie faible admise par les compatibilistes suppose déjà les conditions (c) et (d), l'autonomie ultime les inclut à plus forte raison. Cela suggère qu'il y a au fond une seule condition de la responsabilité morale incluant toutes les autres ; c'est l'autonomie ultime de l'agent lorsqu'il contribue à E.

3.11 Résumé

Dans les chapitres II et III, j'ai défendu les deux principes les plus souvent invoqués par les incompatibilistes : le principe des possibilités alternatives, et le principe du transfert (transfert d'impuissance et de non-responsabilité). Le premier a été critiqué à partir d'exemples souvent fort complexes. Cette complexité les rend difficiles à interpréter et peu efficaces pour mettre en évidence nos intuitions sur la responsabilité morale. J'ai distingué trois catégories d'exemples. Ceux où il y a un intervenant contrefactuel sont réfutés par l'objection du dilemme. Dans d'autres cas, l'agent a une possibilité alternative mais qui ne lui est pas accessible. Ces exemples posent des problèmes de cohérence et sont trop complexes pour que ceux qui les lisent soient convaincus de la responsabilité de l'agent. Enfin, on a proposé des exemples où il y a des possibilités alternatives accessibles à l'agent, mais non pertinentes moralement, or cette donnée exclut intuitivement la responsabilité morale de l'agent.

Si la complexité de ce débat laisse planer un doute sur le principe des possibilités alternatives, l'argument du transfert peut servir à le confirmer. En effet, les exemples les plus convaincants contre le principe des possibilités alternatives (notamment celui de Widerker) supposent ceci :

(AA) Lorsqu'un choix ou une action est l'issue d'un processus indéterministe, celui qui fait ce choix ou cette action le fait *par lui-même*.

Supposons que l'événement x cause entièrement l'action A (de S) de manière indéterministe. Il peut arriver ceci :

1. S n'est pas responsable de x .
2. S n'est pas responsable du fait que x cause A .

Or en vertu de TCNR,

3. S n'est pas responsable de A .

Donc si TCNR est vrai, AA est faux, et les exemples fondés sur AA, comme ceux de Widerker, ne réfutent pas le principe des possibilités alternatives.

Ainsi, le principe du transfert permet de défendre celui des possibilités alternatives. En même temps, il sert à construire l'argument « direct » ou « argument de la conséquence ». C'est cet argument que j'ai défendu dans ce chapitre, en l'appliquant d'abord à l'impuissance (ou à l'inévitabilité) comme ci-dessous :

1. NL
2. NP
3. N(P et L) (en vertu de 1 et 2)
4. N((P et L)→F). Donc :
5. NF (en vertu de 3 et de 4).

Après une brève présentation de cet argument, j'ai défendu les prémisses (1) et (2) (dans les sections 3.2 et 3.3) en défendant les principes de fixité du passé (FP) et des lois de la nature (FL). David Lewis a montré même si nous ne pouvons pas causer un événement violant une loi de la nature, nous pouvons en un sens agir de telle sorte qu'une loi de la nature a été violée à un moment antérieur. Mais cette possibilité ne signifie pas que nous ayons la *capacité* d'agir de la sorte, car la notion de capacité est justement une notion causale. Être capable de faire que E, c'est être capable de produire l'événement E ou d'y contribuer causalement. Ainsi les versions causales de FP et de FL semblent solides, et personne n'a aucun pouvoir sur le passé actuel ou les lois de la nature. Au lieu de contester les principes de fixité (ou les prémisses (1) et (2)), Michael Slote a critiqué l'argument du transfert en soutenant qu'on ne peut pas passer de (1) et (2) à (3). Selon, lui, la modalité exprimée par l'opérateur « N » ne permet pas cette inférence, malgré son air de famille avec la nécessité logique. En effet, « N » se rapproche selon Slote de la non-accidentalité, qui n'est pas agglomérative (elle ne permet pas le passage de (1) et (2) à (3)). Or nous avons vu que Slote comprend la non-accidentalité de telle sorte qu'elle est relative aux intentions ou aux croyances d'une personne. Cette notion est donc très éloignée, contrairement à ce que dit Slote, de l'inévitabilité (« *power-necessity* ») exprimée par l'opérateur N. En effet, même si N peut être indexé sur tel ou tel individu, la notion d'inévitabilité est objective et ne dépend pas

des intentions ou des pensées d'une personne. J'ai donc soutenu que l'objection formulée par Slote n'était pas recevable. L'argument du transfert n'est pas seulement fondé sur des prémisses solides (les propositions (1) et (2)), mais il est valide et concluant.

Dans la section 3.5, j'ai présenté la dernière réponse compatibiliste à l'argument du transfert. Elle consiste à dire que même si cet argument est valide et concluant, il ne prouve pas l'incompatibilité du déterminisme avec la responsabilité morale. Tout ce qu'il montre est l'incompatibilité du déterminisme avec *une certaine* capacité des alternatives. Si P et L entraînent F de manière inévitable, F est inévitable, mais cela prouve-t-il que nous ne sommes pas responsables de F ? Selon Fischer, la responsabilité morale à l'égard d'un événement ne suppose pas le pouvoir d'empêcher cet événement. La conclusion de l'argument du transfert ne concernerait donc pas la responsabilité morale.

Pour écarter cette réponse, il faut construire l'argument du transfert en le faisant porter directement sur la non-responsabilité. On remplace donc l'opérateur « N » par l'opérateur « $\sim R$ ». L'argument s'appuie alors sur le principe du transfert de non-responsabilité (TNR). Mais Fischer, Stump, et Ravizza, ont présenté des contre-exemples à TNR. Après avoir exposé TNR (en 3.7), j'ai discuté deux de ces contre-exemples (en 3.7 et 3.8): un cas de surdétermination *par préemption*, où l'agent est responsable de l'événement en le causant à la place d'un certain facteur également en acte et suffisant pour causer le même événement ; puis un cas de *causalité simultanée* (où deux causes suffisantes séparément pour causer le même événement le causent exactement en même temps). Ces exemples ne nous obligent pas à abandonner le principe du transfert de non-responsabilité. Les trois versions de ce principe que j'ai exposées plus haut (TNR, TCNR, et TCNR*) semblent valables à partir du moment où on considère que les événements sont des particuliers et non des universels.

TNR donne une grande force intuitive à l'idée que la responsabilité morale exige l'autonomie ultime de l'agent. Dans les sections 3.9 et 3.10, j'ai essayé de montrer pourquoi toutes les conditions de la responsabilité se ramènent à l'autonomie ultime : une action ne peut être autonome si elle n'est pas volontaire et consciente, ni si elle est découle du déterminisme causal. L'autonomie ultime inclut donc toutes les autres conditions de la responsabilité morale.

La plupart des philosophes s'accordent sur l'existence d'un lien entre la responsabilité et l'autonomie mais il y a deux conceptions de l'autonomie qui s'affrontent. Les libertariens pensent que l'agent est une cause première. Les compatibilistes, de leur côté, veulent définir l'autonomie de l'agent sans supposer une forme d'indéterminisme. L'autonomie *ultime* implique de l'indéterminisme, mais la nature contient de nombreuses formes d'autonomie compatibles avec un déterminisme strict. Les plantes et les animaux possèdent une forme d'autonomie, puisqu'ils se développent par eux-mêmes, mais ils n'ont pas de responsabilité morale. Quel est donc le degré d'autonomie propre à l'homme et qui en fait un être moralement responsable de ses actes ? J'ai soutenu qu'il pouvait agir avec une autonomie ultime. Pour confirmer cette idée, je vais maintenant critiquer les théories compatibilistes, qui cherchent à énoncer les conditions de la responsabilité morale sans s'appuyer sur elle.

IV L'analyse conditionnelle des pouvoirs et les théories du « moi profond »

4.1 Introduction

Dans les chapitres II et III, j'ai défendu les deux principes incompatibilistes les plus discutés en philosophie analytique : le principe des possibilités alternatives, et celui du transfert d'impuissance. Je pense que les objections contre le premier nous obligent à le reformuler en disant que même si l'agent ne peut pas agir autrement, il doit y avoir *à la source* de son action, des possibilités alternatives qui lui sont accessibles. Cette reformulation nous éloigne d'une compréhension ordinaire du principe PAP, mais semble tenir face à toutes les objections compatibilistes. Le deuxième principe, celui du transfert, corrobore celui du transfert de non-responsabilité (TNR), et tous deux résistent bien aux critiques. Bien que diverses formulations des transferts de capacité, d'impuissance, et de non-responsabilité se heurtent à des contre-exemples, ces principes ne sont pas douteux, car soit les formulations réfutées sont visiblement trop générales et peuvent être aisément précisées, soit les exceptions qu'elles rencontrent sont sans rapport avec le déterminisme causal. Enfin, il est possible de les formuler en les immunisant contre toutes les objections.

Mais si la force intuitive de ces principes n'est pas assez grande pour les compatibilistes, il faut argumenter autrement pour éviter une impasse dialectique¹. Dans les chapitres IV, V, et VI, j'examinerai les théories compatibilistes pour voir si elles atteignent leur objectif : montrer que les conditions de la responsabilité morale ne supposent pas l'existence de possibilités alternatives au sens le plus fort. Pour établir cette conclusion, les compatibilistes doivent élaborer une théorie intelligible des conditions de la responsabilité morale. Il est facile d'énumérer ce qui s'oppose à la responsabilité morale. Par exemple, on n'est pas responsable d'une action si on l'accomplit sous l'effet d'une drogue, d'une pathologie psychique, de

¹ J'emprunte cette expression à J. M. Fischer. Cf. *The Metaphysics of Free Will*, p. 83.

l'hypnose ou d'une manipulation mentale. Pour tenir compte de nouveaux contre-exemples, on peut allonger la liste à volonté en y mentionnant les facteurs qui excluent la responsabilité dans ces contre-exemples. Mais une liste d'exemples ne constitue ni une définition, ni une théorie de la responsabilité morale qui augmente de manière significative notre compréhension de cette notion. Les compatibilistes doivent donc présenter une liste de conditions de la responsabilité morale qui soit valable dans tous les cas possibles, et bien sûr, compatible avec le déterminisme causal. En outre, ces conditions doivent être exprimées avec suffisamment de clarté pour que l'on sache si elles excluent ou non le déterminisme causal.

Dans ce chapitre, j'examinerai d'abord les théories fondées sur l'analyse conditionnelle du verbe « pouvoir ». Les philosophes déterministes ou compatibilistes ont souvent défini la liberté de façon conditionnelle : comme la faculté de faire *x* si on le veut. Cette définition exprime une conception de l'autonomie propre à l'homme : il est autonome au sens où il a la faculté d'agir selon sa propre volonté. Nous verrons que cette idée, lorsqu'elle est utilisée pour définir la liberté non seulement dans l'action, mais aussi dans les choix ou volontés, implique une hiérarchie de désirs comprenant des désirs de second ordre par lesquels l'agent contrôle ses désirs de premier ordre. Cette idée de hiérarchie des désirs défendue par Frankfurt, Lehrer, Watson et Taylor, permet une conception plus élaborée de l'autonomie et du contrôle de soi. Néanmoins, nous verrons qu'elle se heurte à plusieurs difficultés qui nous obligent à comprendre autrement l'autonomie de l'agent. J'examinerai ensuite une deuxième stratégie compatibiliste permettant de définir l'autonomie de l'agent sans aller jusqu'à admettre une autonomie *ultime* (incompatible avec le déterminisme) : elle consiste à dire que l'agent est autonome lorsqu'il possède une raison saine et suffisamment développée. C'est la stratégie de Suzan Wolf.

Dans le chapitre V, je soutiendrai que les théories compatibilistes n'ont pas les moyens de rendre compte des cas d'omissions, car lorsqu'un agent est responsable d'un événement par omission, sa responsabilité suppose qu'il pouvait agir de manière à empêcher cet événement. J'exposerai une théorie libertarienne de la responsabilité des omissions, puis je soutiendrai qu'il n'y a pas de réponse satisfaisante, du point de vue compatibiliste, pour rendre compte des cas d'omissions. Je critiquerai les trois solutions compatibilistes : la réduction des omissions

aux actions, l'analyse conditionnelle du verbe « pouvoir », et l'analyse des pouvoirs en termes de mondes possibles.

Dans le chapitre VI, je soutiendrai que le compatibilisme a besoin de l'analyse des pouvoirs en termes de mondes possibles, et que cette solution pose un problème de clarté insurmontable.

4.2 La définition conditionnelle de la liberté

Dans le premier chapitre², nous avons vu que certaines facultés pouvaient se définir de façon conditionnelle, par exemple les dispositions involontaires appartenant aux plantes, aux animaux, et même aux substances inertes (la faculté de croître, de digérer, la solubilité du sucre ou la faculté de l'aimant d'attirer le fer). C'est également le cas d'autres puissances qui ne passent à l'acte que sous l'effet de la volonté. Dire qu'une substance est soluble dans l'eau (capacité du premier genre) signifie que si elle est plongée dans l'eau, elle se dissout ; de la même façon, « Pierre est capable de nager » (capacité du deuxième genre) signifie que s'il veut nager (et en a l'occasion), il nage.

Puisque la liberté est également une faculté, il est naturel de supposer qu'elle pourra se définir de la même façon, ce qui donnerait un énoncé de la forme :

(déf.) A est libre de faire $x \leftrightarrow$ si A veut faire x , il fait x .

Dans *l'Essai sur l'Entendement humain*, Locke admet une définition de ce genre. Puisque sa position est caractéristique, je la présenterai de façon développée avant de passer à d'autres théories plus récentes. D'abord, il imagine cet exemple :

Supposez un homme transporté pendant son sommeil dans une chambre où se trouve une personne qu'il est impatient de voir et qu'il y soit enfermé de sorte qu'il soit hors de son pouvoir de sortir ; il se réveille, il est heureux de se trouver en compagnie si désirée et il demeure volontairement là, c'est-à-dire il préfère

² Cf. *supra*, p. 36.

rester plutôt que s'en aller. Ma question : n'est-ce pas rester volontairement ? Je pense que personne n'en doutera ; et pourtant, étant enfermé, il n'a évidemment pas la liberté de ne pas rester, il n'a aucune liberté de sortir. Ainsi, *la liberté n'est pas une idée attachée à la volition* ou à la préférence, mais à la personne qui a le pouvoir de faire ou d'éviter de faire selon que l'esprit choisira ou ordonnera. Notre idée de liberté a la même extension que ce pouvoir et pas plus. Car là où une limite vient s'opposer à ce pouvoir, là où une contrainte ôte l'indifférence ou la capacité d'agir en l'un ou l'autre sens, la liberté disparaît aussitôt et avec elle la notion que l'on en a³.

En somme, Locke distingue deux définitions. Selon lui, la liberté n'est pas la capacité de faire ce qu'on veut ; il faut définir la *liberté de faire x* comme la capacité de faire *x si on le veut*. L'individu imaginé par Locke fait ce qu'il veut puisqu'il veut rester en compagnie de sa bien-aimée. Pourtant il n'est pas libre car s'il voulait sortir, il ne le pourrait pas. Locke distingue ainsi l'action *libre* et l'action *volontaire*. La deuxième définition, comme la première, établit un lien entre la liberté et la volonté, mais cette fois, il ne s'agit plus seulement de ce que *veut* l'agent, mais de ce qu'il *peut vouloir*. La définition conditionnelle de la liberté est donc liée à l'idée que l'agent pourrait agir autrement. Faut-il parler ici de possibilités alternatives ? Oui et non. Il ne s'agit pas des possibilités alternatives au sens strict, à moins de préciser que l'agent a la possibilité d'agir autrement *compte tenu des lois de la nature et du passé actuel*⁴. Tant qu'on ne précise pas de quelles possibilités il s'agit, la définition est parfaitement compatible avec la thèse déterministe.

En ce qui concerne Locke, sa position n'apparaît pas clairement, d'emblée, comme une position déterministe ou même compatibiliste. On remarque néanmoins qu'il veut concilier l'idée d'un pouvoir des contraires avec le principe métaphysique selon lequel rien n'est sans cause. Sa thèse est donc bien compatibiliste. Au paragraphe 13, il définit la nécessité comme étant soit la « compulsion », soit « l'inhibition » selon les cas⁵. Une définition aussi restrictive s'accorde bien avec le compatibilisme, car elle refuse d'appeler « nécessité » une nécessité causale interne à l'agent, par laquelle il agirait en vertu de son caractère, son naturel ou son tempérament.

³ John Locke, *Essai sur l'Entendement humain*, trad. Jean-Michel Vienne, II, 21, § 10, p. 381.

⁴ cf. *supra*, p.23.

⁵ John Locke, *Essai sur l'Entendement humain*, p.383.

L'interprétation de la thèse de Locke est rendue difficile par le fait qu'il a corrigé plusieurs fois son texte entre la première et la cinquième édition de l'*Essai*⁶. Au début, sa position était simple et franche, mais il s'est aperçu qu'il devait accorder une plus grande place à la liberté de la pensée dans sa psychologie morale, ce qui l'a amené à introduire des changements concernant la suspension du désir, et la liberté par rapport au vouloir.

Au paragraphe 25, Locke affirme que la liberté appartient à l'agent et non à la volonté de l'agent. L'expression anglaise « *free will* » (volonté libre) serait donc fallacieuse. Quand on agit librement, on a la possibilité d'agir autrement ou de ne pas agir du tout, mais on n'a ni la possibilité de vouloir autrement, ni celle de ne rien vouloir :

Demander si on est libre de vouloir soit le mouvement soit le repos, soit parler soit le silence, bref ce qui plaît, c'est demander si on peut *vouloir* ce qu'on veut, ou se plaire à ce qui plaît : une question qui à mon sens n'a pas besoin de réponse. Ceux qui en font un problème doivent supposer une volonté qui déterminerait les actes d'une autre, et une autre qui déterminerait celle-là, et ainsi de suite à l'infini⁷.

Selon Locke, la liberté supposerait donc la capacité d'agir autrement, mais non celle de vouloir autrement. Si cette définition est correcte, il est certain que la liberté est compatible avec le déterminisme, mais est-elle correcte, et Locke lui-même est-il déterministe ? Dans les paragraphes suivants, il se demande ce qui détermine la volonté. Sa réponse est que toutes les fois où la volonté de l'agent se porte vers un nouvel objet, c'est une sorte de malaise (plus ou moins équivalent au désir) qui l'y entraîne. Plus loin, il nuance cette affirmation car l'expérience nous montre parfois que

l'esprit a le pouvoir de *suspendre* l'exécution et la satisfaction de n'importe quel désir et donc de chacun à son tour ; il a ainsi la liberté d'en étudier l'objet, de les examiner sous toutes leurs faces, de les comparer à d'autres. [...]

⁶ Sur l'évolution de la position de Locke, cf. Richard Glauser, « Thinking and Willing in Locke's Theory of Human Freedom », in *Dialogue* XLII (2003), p.695-724. L'auteur de cet article soutient que malgré ces retouches successives, la position finale de Locke dans l'*Essai* est cohérente.

⁷ John Locke, *Essai sur l'Entendement humain*, p. 392.

Ceci me semble être la source de toute liberté, et ce en quoi paraît consister ce que l'on appelle (à tort, à mon sens) volonté libre⁸.

Locke suppose donc que nous avons la possibilité d'agir autrement en suspendant ou en retardant l'exécution d'un désir. Est-ce à dire que nous avons des possibilités alternatives au sens fort et que le déterminisme est faux ? Non, car lorsque notre vouloir n'est pas déterminé par le désir (ou le malaise), il l'est par notre jugement, c'est-à-dire par notre raison, ou par *les* raisons⁹. Et comme Leibniz¹⁰, Locke défend cette idée en évoquant Dieu et les créatures spirituelles supérieures à l'homme : ils sont libres alors même qu'ils sont « fermement déterminés dans le choix du bien »¹¹. Sur ce point, Locke et Leibniz sont donc d'accord : être libre, c'est être déterminé par les raisons. Locke conclut et résume sa position dans le paragraphe 56 :

La liberté consiste, c'est évident, dans le pouvoir de faire ou de ne pas faire ; de faire ou de s'abstenir de faire comme on le *veut* ; on ne peut le nier. Mais ceci ne semble concerner que les actions de l'homme qui suivent une volition, d'où l'on pose cette nouvelle question : a-t-il ou non la liberté de *vouloir* ? On y a répondu que dans la plupart des cas, un homme n'a pas la liberté de s'abstenir d'un acte de volition : il doit exécuter un acte de sa *volonté*, qui fait exister ou ne pas exister l'action proposée.

Il existe pourtant un cas où l'homme dispose de liberté à l'égard de sa *volonté* : c'est quand il choisit un bien éloigné comme fin à poursuivre : ici, un homme peut suspendre la détermination de son choix pour ou contre la chose proposée, jusqu'à ce qu'il ait examiné si ce bien est effectivement de nature, en lui-même et par ses conséquences, à lui apporter le bonheur. [...]

On peut ainsi voir comment il se fait qu'un homme puisse à juste titre mériter une punition, même s'il est certain que, dans toutes les actions qu'il *veut*, il le veuille et veuille nécessairement ce qu'il juge alors être un bien. Car bien que sa volonté soit toujours déterminée par ce que son entendement juge bon, cela ne

⁸ *Ibid.*, p. 413.

⁹ Pour Locke, notre volonté est déterminée ordinairement par nos jugements en même temps que par le malaise. Voir § 46, n°263 (p.412) sur le malaise, et § 48, n°265 (p.414-415) sur le jugement. Cette idée n'est pas contradictoire ; elle revient à dire ceci : nous ne voulons que ce que nous croyons être bon pour nous, et nous le voulons parce que son absence provoque en nous un malaise.

¹⁰ Par exemple, dans la *Conversation sur la liberté et le destin*. Cf. *infra*, p.316.

¹¹ *Ibid.*, § 49 (p. 415).

l'excuse pourtant pas : par un choix trop hâtif de son propre chef, il s'est imposé à lui-même des mesures erronées de bien et de mal » [...] l'égarement qui s'en suit doit être imputé à son propre choix¹².

Ainsi, selon Locke, nous n'avons pas de *libre arbitre (free will)*, car la volonté n'est pas libre, mais nous possédons parfois une sorte de *liberté par rapport au vouloir* : celle de suspendre notre jugement et le vouloir qui en découle en prenant le temps d'examiner quel est le bien le plus grand. Lorsque nous voyons enfin quel est ce bien, nous ressentons un malaise en son absence, lequel malaise nous détermine à le vouloir. Une question demeure : cette faculté de choisir autrement liée à la suspension du jugement implique-t-elle de la contingence dans les actions humaines ? Si le choix de l'agent (lorsqu'il arrête son esprit sur certaines raisons ou l'en détourne) est nécessaire en vertu du passé actuel et des lois de la nature, il est déterminé causalement. Puisque c'est probablement l'opinion de Locke, nous pouvons penser que dans son chapitre sur les pouvoirs, la définition conditionnelle de la liberté lui sert à défendre une position compatibiliste sur le déterminisme et la responsabilité morale. En dernière analyse, Locke admet que la responsabilité morale suppose une capacité de vouloir autrement découlant d'une capacité de fixer son attention sur les raisons ou au contraire de l'en détourner. Mais cette liberté dans les jugements qui déterminent la volonté n'est pas ce qu'on appelle habituellement le libre arbitre. Pour Locke, la liberté s'applique aux actions mais non aux volontés (c'est l'action – ou l'agent – qui est libre, et non la volonté). Mais surtout, Locke admet le principe selon lequel rien n'est sans cause, et il pense qu'on est entraîné dans une régression à l'infini à partir du moment où on applique la définition conditionnelle à la volonté elle-même. Voici en effet ce que dirait cette définition :

(déf.) S veut librement $x \leftrightarrow$ Si S veut vouloir x , il veut x .

Cette définition est en effet étrange. Mais Locke pourrait envisager des situations où nous avons des volontés ou des désirs *de second ordre*, par exemple une personne qui veut en finir avec son attirance pour le tabac. C'est parce qu'il n'envisage pas ce genre de volontés et craint une régression à l'infini qu'il refuse d'appliquer sa définition conditionnelle de la liberté à autre chose que les actions. Ce refus ne l'empêche pas de croire que nous avons la capacité de vouloir autrement, seulement cette faculté n'est pas liée pour lui à des volitions de

¹² *Ibid.*, p. 423.

second ordre par lesquelles nous rejetons ou adoptons des volitions de premier ordre. Elle est plutôt liée à celle d'arrêter ou de prolonger la considération des raisons en faveur de tel ou tel choix¹³.

4.3 La définition conditionnelle redoublée. Les niveaux de désirs

La définition conditionnelle de la liberté permet d'accorder une place à l'idée de possibilités alternatives. Elle est ainsi plus proche de nos croyances spontanées que la définition de la liberté comme le simple fait d'accomplir quelque chose volontairement. C'est ce qu'a bien vu Locke dans *l'Essai sur l'Entendement humain*. Pourtant, nous allons voir que la définition conditionnelle est insuffisante en elle-même. Soit il faut l'abandonner, soit il faut la compléter. Une manière de la compléter est de la redoubler en l'appliquant à la fois à

¹³ Chez Moore, nous trouvons également une définition conditionnelle de la liberté. cf. G. E. Moore, "Free Will" (à l'origine, ch. VI de *Ethics*, William & Norgate, London, 1912, trad. fr. par Michel Gouverneur dans G. E. Moore, *Principia Ethica*, PUF, Paris, 1998, p. 329-343). La position de Moore est compatibiliste :

[...] en réalité, il est extrêmement douteux que le libre arbitre soit d'une quelconque façon en contradiction avec le principe selon lequel tout a une cause. Mais que cette contradiction existe ou non dépend d'une question très délicate sur le sens du terme « pouvoir ». Tout ce qu'il y a de certain sur cette question, c'est 1/ que, si nous avons le libre arbitre, il est probablement vrai qu'en *un certain* sens nous aurions parfois *pu* faire ce que nous n'avons pas fait effectivement ; et 2/ que, si tout a une cause, il est probablement vrai qu'en *un certain* sens nous *n'aurions jamais pu* faire ce que nous n'avons pas fait effectivement. Ce qui est en revanche tout à fait incertain, et ce sur quoi il est encore nécessaire de mener nos investigations, c'est de savoir si ces deux sens du mot « pouvoir » sont identiques. (Ibid., p. 337)

Pour Moore, la définition conditionnelle de la liberté est un moyen de concilier ces deux intuitions, à savoir que nous pouvions agir autrement, mais qu'en un sens nous ne le pouvions pas. En d'autres termes, la première intuition est que la responsabilité morale suppose la capacité d'agir autrement, et la deuxième est que tout a une cause (principe de causalité). Ces deux idées indiquent la solution, car pour les concilier, il est nécessaire d'employer deux sens différents du mot « pouvoir » :

En quel sens *aurais-je pu* faire, par exemple, 2 km de marche en vingt minutes ce matin, bien que je ne l'aie pas fait ? La réponse qui se présente à l'esprit de façon tout à fait évidente, c'est qu'après tout je veux peut-être simplement dire que j'aurais pu le faire, *si* tel avait été mon choix ; autrement dit, c'est que (pour éviter une possible complication) nous ferions peut-être mieux de dire : « je l'*aurais* fait *si* j'avais choisi de le faire ». (Ibid., p. 337, traduction légèrement modifiée).

Mais à la différence de Locke, Moore admet ensuite que nous n'avons pas seulement le pouvoir d'agir autrement, mais aussi celui de vouloir autrement. Il le définit également de façon conditionnelle : nous avons le pouvoir de vouloir autrement *si nous le voulons*. cf. p.341 : « Je pense qu'il est souvent vrai, indubitablement, que nous aurions choisi de faire une chose précise si nous avions choisi de faire le choix, souvent vrai aussi que s'il est très souvent en notre pouvoir de faire un choix, c'est en ce sens primordial de l'expression. Très certainement, faire un effort pour nous amener à *choisir* une ligne d'action, cela se fait ; et je pense sans aucun doute que, bien des fois, si nous *avons* fait un tel effort, nous aurions fait un choix qu'en réalité nous n'avons pas fait ».

l'action libre et à la volonté libre. Cela conduit à définir la liberté comme *accord des désirs de premier niveau avec les désirs de niveau supérieur* (par exemple, on ne fume pas librement si on fait ce qu'on désire mais qu'on ne *veut pas* désirer fumer). J'examinerai les différents problèmes que pose cette conception de la liberté. Tout d'abord, il faut se demander pourquoi l'existence d'un second niveau de désir permet la liberté. Ensuite, l'existence d'un second niveau est-elle suffisante, ou bien faut-il imaginer des désirs de niveaux 3, 4, 5, etc., voire même une infinité de niveaux ? Ces questions nous amèneront finalement à considérer la relation entre le contrôle de soi-même et l'identité avec soi-même.

Revenons d'abord à la définition conditionnelle simple. Elle pose que :

(déf.) S est libre de faire A $\leftrightarrow$ Si S veut faire A, il fait A.

Ici, on ne fait pas de différence entre vouloir et désirer. Or on sait que pour être libre, il ne suffit pas d'agir selon ses désirs. Il est souvent nécessaire de lutter contre eux, de les réprimer ou de les refréner, pour atteindre une liberté complète. L'agent lui-même peut désapprouver ses désirs. Dans ce cas, agir selon ses désirs serait pour lui une forme d'aliénation. Par exemple, une personne qui lutte contre ses penchants alcooliques n'est libre que si elle évite de les satisfaire. Ce constat amène tout naturellement à définir la liberté comme le fait, non seulement d'agir selon ses désirs, mais de les contrôler, c'est-à-dire d'avoir des désirs que l'on approuve soi-même. Harry Frankfurt a lancé l'idée que le libre arbitre était la faculté d'avoir des désirs de second niveau (désirs *au sujet de* ses désirs). Les animaux possèdent la conscience à un certain degré, mais n'ont pas de désirs de second niveau, comme le désir de désirer travailler, ou le désir de ne pas désirer fumer. La liberté au sens plein du mot suppose que l'agent ait des désirs de second niveau en accord avec ses désirs de premier niveau. Lorsqu'il y a un conflit non résolu entre ces deux niveaux d'inclination, l'agent ne fait pas vraiment ce qu'il veut. C'est la situation décrite par St Paul : « Vraiment, ce que je fais, je ne le comprends pas, car je ne fais pas ce que je veux, mais je fais ce que je hais. [...] Or si je fais ce que je ne veux pas, ce n'est plus moi qui accomplit l'action, mais le péché qui habite en moi. »¹⁴

¹⁴ *Épître aux Romains* 7, 15 et 20, traduction de la Bible de Jérusalem.

La définition conditionnelle simple est donc insuffisante parce qu'elle néglige le fait que la liberté suppose le contrôle de ses propres désirs. Selon Keith Lehrer, ce défaut ne prouve pas qu'elle soit à rejeter entièrement, mais seulement qu'elle n'indique pas l'ensemble de conditions à la fois nécessaires et suffisantes pour qu'un agent soit libre. Il faudrait donc la compléter en énonçant des conditions supplémentaires. La première manière de s'y prendre - la plus simple et la moins coûteuse - est sans doute de redoubler la définition conditionnelle en l'appliquant non seulement aux actions, mais aux volontés. À la première condition, on ajoutera une deuxième, également de forme conditionnelle :

(déf.) S est libre de faire A $\leftrightarrow$ 1°) Si S veut faire A, il fait A, 2°) Si S préfère ne pas vouloir faire A, il ne veut pas faire A.

Cette définition indique que pour être libre, l'agent doit contrôler les désirs ou les choix qui causent ses actions. Il a le pouvoir de vouloir autrement. Donc sa liberté consiste à la fois dans le contrôle de ses actions et dans le contrôle de ses inclinations. L'existence du second niveau de désirs permet à l'agent de réviser ses désirs ou d'exercer sur eux un contrôle conscient¹⁵. Rappelons cependant qu'une action ou une omission peut être volontaire sans être voulue *consciemment* (ou sans être *actuellement consciente*). L'agent doit seulement avoir la capacité de s'apercevoir de ce qu'il fait, et il en va de même pour les désirs¹⁶.

La première difficulté se présente ainsi : est-il suffisant que le désir soit redoublé ? Ne pouvons-nous imaginer des situations où le désir de niveau 2 (appelons-le « désir 2 ») soit lui-même produit par une cause extérieure, par exemple un savant fou manipulant le cerveau de l'agent ? Dans ce cas, même s'il y a un parfait accord entre les désirs des niveaux 1 et 2, on

¹⁵ Dans son article sur la notion de personne, Harry Frankfurt déclare que les désirs de second niveau sont propres à l'homme. Il précise que ces désirs ne sont pas nécessairement conscients : « J'interpréterai les énoncés de la forme « A désire faire X » comme couvrant un éventail assez large de situations. Ils peuvent être vrais même lorsque [...] A n'a pas conscience d'un quelconque sentiment se rapportant à l'accomplissement de X, lorsqu'il ne se rend pas compte qu'il désire faire X, lorsqu'il se trompe à propos de ce qu'il désire, et croit qu'il ne désire pas faire X, lorsque d'autres désirs entrent en conflit avec son désir de faire X, ou lorsque son désir est ambivalent. Les désirs en question peuvent donc être conscients ou inconscients, ils ne doivent pas nécessairement être univoques et A peut se tromper à leur propos. » (H. Frankfurt, « La liberté de la volonté et la notion de personne », in *Théorie de l'action*, anthologie de textes de philosophie analytique sur l'action réunis par Marc Neuberg, éd. Mardaga, Paris, 1995)

¹⁶ Sur ce point, la thèse de Frankfurt n'est pas différente de la mienne.

doute que l'agent soit réellement libre. Supposons qu'il désire fumer, mais désire résister à ce désir. Pour arrêter de fumer, il essaie tout, même l'acupuncture, mais rien ne semble efficace. Un jour, ce conflit d'inclinations se résout grâce à un savant fou exerçant, à l'insu de l'agent, une influence sur ses désirs de niveau 2. Mais ce qu'il produit en lui, c'est le désir de désirer fumer. Désormais, l'agent a donc des désirs 2 en accord avec ses désirs 1, mais contrairement à la définition proposée, il n'est pas vraiment libre. En effet, nous ne pouvons pas le considérer comme responsable de son comportement. Bien au contraire, il est tombé dans une aliénation beaucoup plus profonde qu'auparavant.

Pourquoi donc accorder une telle importance à l'existence d'un deuxième niveau d'inclinations ? N'est-ce pas seulement un moyen de repousser le problème ? Après tout, s'il y a un troisième niveau, rien n'empêche d'imaginer des situations où les désirs de ce niveau sont produits par une cause extérieure échappant au contrôle de l'agent. Et ainsi de suite, car cela vaut également pour les niveaux 4, 5, 6, etc. La seule solution serait de dire que le nombre de niveaux est infini. Ainsi, chaque niveau serait contrôlé par un niveau supérieur, alors que si le nombre de niveaux est fini, le plus élevé dans cette hiérarchie est nécessairement un niveau qui n'est pas l'objet d'un désir conscient.

La thèse de Frankfurt est liée à une certaine idée du contrôle exercé par l'agent sur ses propres désirs : c'est une activité consciente, qui approuve ou rejette les désirs de premier niveau. Il est en effet tentant de supposer que pour contrôler ses désirs, l'agent ait besoin de les viser consciemment. Sans cela, on ne voit pas comment il pourrait les juger bons ou mauvais, puis les favoriser ou les réprimer. Ce qui échappe à la conscience de l'agent échappe à son contrôle, c'est pourquoi il n'est responsable que des désirs dont il peut être conscient. Cette idée expliquerait en même temps pourquoi nous n'attribuons pas de responsabilité morale aux êtres vivants dépourvus de désirs de second niveau.

Néanmoins, elle entraîne plusieurs difficultés. Premièrement, s'il doit y avoir dans l'esprit de l'agent une infinité de niveaux de conscience, cela ne suppose-t-il pas une capacité d'attention proprement surhumaine ? Sommes-nous vraiment capables de penser consciemment à un désir de désirer désirer (etc.) quelque chose ? Deuxièmement, certains

philosophes comme Gary Watson pensent que la pluralité des niveaux n'ajoute rien tant qu'ils ne sont pas essentiellement différents. Concentrons-nous d'abord sur cette objection.

Frankfurt estime que les inclinations de second niveau ne sont pas des volontés (ou volitions) ; elles sont des désirs, et parmi les désirs de premier niveau, seuls ceux qui l'emportent sont aussi des volontés. Ainsi, toute inclination est un désir, et on appelle « volonté » un désir que l'agent choisit de réaliser, voire un désir efficient¹⁷.

Ici, la volonté se trouve donc toute entière au premier niveau, ce qui n'est pas évident, car on pourrait supposer qu'elle exerce un contrôle conscient sur les désirs. Gary Watson¹⁸ a proposé une théorie du même genre que celle de Frankfurt, mais pour lui, il y a au premier niveau les « simples désirs », et au second niveau, les désirs qui sont des valeurs (ou qui impliquent une évaluation). La différence entre ces deux types de désirs n'est pas liée à la forme logique de leur description, mais à une différence d'origine : certains désirs ne sont que des appétits sensibles, ou des réponses conditionnées à des *stimuli*, tandis que d'autres expriment des jugements de l'agent lui-même sur la valeur des choses qu'il désire. Ces derniers désirs proviennent donc de ce que Watson nomme le « système d'évaluations » de l'agent.

Watson oppose deux visions différentes des conflits d'inclinations : celle de Hume et celle de Platon. Hume conçoit la raison comme une faculté dont la fonction est purement épistémique ; c'est la faculté de raisonner, de calculer, d'argumenter. Elle ne constitue pas une source autonome de motivations ou d'inclinations puisque son rôle est seulement de distinguer le vrai du faux dans les questions de faits (« *matters of fact* ») et dans les relations entre les idées. Dans le domaine pratique, elle évalue la probabilité d'obtenir tel ou tel bien et de concilier les désirs entre eux, mais cette activité est encore une forme de calcul. Pour

¹⁷ « On identifie la volonté d'un agent en identifiant le ou les désirs qui sont soit les motifs de son action présente, soit les motifs de son action à venir, soit ses motifs au cas où il agirait. La volonté d'un agent est donc identique à un ou plusieurs de ses désirs de premier niveau. Mais cette notion de volonté n'a pas la même extension que la notion de désirs de premier niveau, en effet, elle ne comprend pas toutes les propensions, plus ou moins fortes, à agir, mais seulement les désirs efficients, ceux qui amènent (ou amèneront, ou amèneraient) la personne à passer à l'acte. »

¹⁸ Gary Watson, "Free Agency", *Journal of Philosophy* LXXII (1975), p. 205-220, reproduit dans *Free Will*, anthologie d'articles sur le libre arbitre réunis par Gary Watson, Oxford University Press, Oxford, 2003.

Platon, la raison est au contraire une partie de l'âme qui a ses propres inclinations. Notre âme se divise en deux parties, une partie rationnelle et une partie irrationnelle, chacune ayant ses propres tendances, ce qui donne lieu à des conflits ou à des tensions. Dans un passage du *Phèdre*¹⁹, Platon compare notre âme à un attelage mal coordonné dont les chevaux ne se dirigent pas dans la même direction. La partie rationnelle de l'âme est attirée par le Bien, alors que la partie non-rationnelle tend vers les réalités matérielles et sensibles. Comme l'écrit Watson, « Platon attire l'attention sur le fait que c'est une chose de penser qu'un état de choses est bon, est louable, ou mérite d'être recherché, et une autre de désirer ou de vouloir simplement que cet état de choses se réalise »²⁰. On peut en effet désirer un état de chose sans l'évaluer à l'aide de concepts du Bien et du Mal. Dans beaucoup d'autres philosophies, comme chez Aristote et Thomas d'Aquin, il y a également deux niveaux d'inclinations liés à des modes de connaissance différents. Les désirs sont, au moins pour une part, liés à la sensation et aux plaisirs des sens, tandis que la volonté est liée à la connaissance conceptuelle. Seuls les hommes ont une volonté parce qu'eux seuls ont une intelligence, qui permet d'évaluer les désirs.

De plus, d'après Watson, « il ne peut y avoir de problème de l'action libre que si la conception platonicienne de l'âme est (à peu près) correcte »²¹. En effet, cette conception répond à la question : « en quel sens ce qu'on veut le plus peut-il être différent de l'objet du désir le plus fort ? ». L'existence de conflits d'inclinations est liée au fait que nos désirs viennent de plusieurs sources différentes. Il y a plusieurs causes de conflits d'inclinations, mais ceux qui n'existent que chez les êtres auxquels nous attribuons le libre arbitre ne sont pas seulement des conflits entre plusieurs désirs sensibles. L'âne de Buridan est partagé entre deux désirs de même nature, ou deux manières différentes de réaliser le même désir. Il a soif, et il est attiré par deux lieux à la fois. Chez les animaux, il y a possibilité de conflits entre des désirs sensibles de natures différentes, par exemple la faim et le désir du repos. Nous

¹⁹ Platon, *Phèdre*, 246a-c. cf. aussi 237e : « Il faut savoir qu'il y a dans chacun de nous deux principes qui nous gouvernent et nous dirigent et que nous suivons où ils nous mènent : l'un est le désir inné du plaisir, l'autre l'idée acquise qu'il faut rechercher le bien. Ces deux principes tantôt s'accordent, tantôt se combattent en nous, et tantôt c'est l'un, tantôt c'est l'autre qui triomphe ». (traduction Émile Chambry, GF-Flammarion)

²⁰ Gary Watson, "Free Agency", p. 340 dans Watson (2003).

²¹ *Ibid.*, p. 341.

supposons que le désir le plus fort est celui qui va l'emporter. Le choix de l'animal répondrait ainsi à une mécanique des désirs dans laquelle l'inclination la plus forte serait nécessairement celle qui l'emporte à la fin (nous supposons donc que si une inclination l'emporte, c'est parce qu'elle était la plus forte). Mais chez l'homme, il y a des conflits d'un autre ordre : entre des désirs et des volontés liées à des évaluations. Il y a donc en lui une autre source de tendances que la sensation. Cette source est la raison ou l'intellect, en tout cas la faculté d'évaluer les choses à partir d'un concept du bien et du mal. Elle est liée à la connaissance conceptuelle, et non à une connaissance purement sensible²².

Ce qui marque la différence entre les deux niveaux de désirs, c'est que nos désirs de premier niveau existent indépendamment de ce que nous en pensons. Par exemple, le désir sexuel, le désir de manger ou de boire, s'imposent d'abord à l'agent, et c'est seulement ensuite qu'il peut intervenir en les atténuant ou en les réprimant, ce qui suppose un effort. Il n'atteint la liberté au sens plein du mot que lorsque sa structure motivationnelle (incluant tous ses désirs) coïncide avec sa structure de valeurs²³.

La théorie de Watson ajoute bien quelque chose à l'idée première de Frankfurt, mais ne résout pas toutes les difficultés. Il importe en particulier de savoir si le deuxième niveau d'inclinations est suffisant ou s'il faut imaginer des niveaux supérieurs, voire une infinité de niveaux. Watson observe que les jugements pratiques sont liés aux désirs de second niveau, et

²² Chez Thomas d'Aquin, la volonté est un appétit lié à la connaissance conceptuelle du Bien. Eleonore Stump applique cette conception de la volonté au débat sur la hiérarchie des désirs. cf. Eleonore Stump, « Intellect, Will, and Alternate Possibilities » (1990), in Fischer et Ravizza (1993). Elle résume cette conception p.251 : « La volonté est pour St Thomas une bouche aveugle. C'est une faim de bonté, mais par « bonté », l'Aquinat entend ici la bonté en général et non telle ou telle bonne chose particulière ; autrement dit, la volonté veut ce qui est bon, l'expression « ce qui est bon » étant à comprendre ici de façon attributive et non référentielle. Et par elle-même, la volonté ne fournit aucune détermination de la bonté. L'acte d'appréhender ou de juger les choses du point de vue de la bonté est l'affaire de l'intellect. L'intellect présente certaines choses à la volonté comme étant bonnes sous certaines descriptions, et la volonté les veut parce qu'elle est un appétit pour le bien et qu'elles sont appréhendées comme bonnes. Pour cette raison, on dit que l'intellect meut la volonté, non à la manière d'une cause efficiente, mais à la manière d'une cause finale, parce que ce qui est compris comme bon meut la volonté en tant que fin (ST Ia q.82 a.4). »

²³ Les valeurs sont donc, pour Watson, des conceptions de ce qui est bon ou désirable. Elles viennent de notre intelligence conceptuelle. Et ce sont elles que nous invoquons en dernière analyse lorsque nous donnons nos raisons d'agir de telle ou telle manière.

que l'existence de niveaux supérieurs est douteuse, mais il ne traite pas à fond cette difficulté. Il conclut en disant que sa propre théorie repose moins sur la distinction entre des *niveaux* différents que sur la distinction entre plusieurs sources indépendantes de motivations²⁴. Cependant, n'est-il pas nécessaire qu'il y ait des volontés de plusieurs niveaux ? Si les jugements de valeur permettant à l'agent de contrôler ses désirs de premier niveau ne sont pas eux-mêmes voulus ou approuvés par l'agent, comment pouvons-nous dire qu'il les contrôle ? N'est-il pas nécessaire qu'il y ait une infinité de niveaux dans les volontés de l'agent ? Sans cela, le dernier niveau échapperait à son contrôle. Imaginons un adolescent iroquois qui, pendant les rites d'initiations, maîtrise courageusement ses désirs sensibles. Si cela est entièrement dû aux valeurs iroquoises transmises par l'éducation et admises par lui de façon inconsciente, pouvons-nous dire qu'il est responsable de son comportement ? S'il n'avait ni le pouvoir d'acquiescer d'autres valeurs, ni celui d'agir en contradiction avec les valeurs de son ethnie, il ne semble pas responsable de ce qu'il fait. Un homme ne s'identifie pas purement et simplement avec son système de valeurs ; il faut qu'il le contrôle pour s'identifier en quelque sorte avec lui.

4.4 Y a-t-il une infinité de niveaux dans les inclinations ?

L'argument que nous venons d'évoquer milite en faveur d'une infinité de niveaux. Mais cette idée paraît elle aussi invraisemblable, car nous ne sommes pas capables de nous représenter simultanément une infinité de désirs. On a souvent dit qu'il était psychologiquement irréaliste d'imaginer une hiérarchie de désirs comprenant une infinité de niveaux : des désirs de second ordre, puis de niveaux 3, 4, 5, 6, etc. Nous n'avons pas l'impression d'avoir des désirs aussi abstraits, ni aussi complexes.

Commençons par une observation importante : il y a des conflits de désirs entre les niveaux 1 et 2, mais il n'y en a guère entre les niveaux supérieurs. Pour les niveaux 2 et 3, cette affirmation mérite d'être nuancée, mais on imagine difficilement une opposition entre les niveaux 5 et 6, par exemple. Or pourquoi serait-ce impossible ? Bien entendu, il n'y a pas

²⁴ Cf. Gary Watson, "Free Agency", p. 350 dans Watson (2003).

ici d'impossibilité *conceptuelle*, car nous *concevons* aisément une divergence entre le désir 5 d'une barre de chocolat, et le désir 6 (désir de ne pas désirer-5 la manger). Nous la *concevons*, mais nous ne *l'imaginons* pas. Autrement dit, nous voyons bien qu'il y a une différence conceptuelle entre ce conflit et le même conflit transposé un niveau plus bas, mais nous avons bien du mal à nous représenter une situation où ce conflit existerait. Je pense que cela tient à des raisons psychologiques.

Pour examiner ce point, j'utiliserai un passage de la *Somme Théologique* de St Thomas d'Aquin. Dans la question 86 de la *prima pars*, il se demande si notre intellect peut connaître des infinis. Dans la réponse, il distingue trois sortes de connaissances : on peut connaître en acte, en puissance, et en habitus :

[...] dans les réalités matérielles, il n'y a pas d'infini en acte, mais seulement un infini en puissance, en tant que l'une succède à l'autre, d'après Aristote. Par conséquent on trouve dans notre intelligence un infini en puissance, en tant qu'elle considère un objet après un autre. Car notre intelligence ne connaît jamais tant de choses qu'elle n'en puisse connaître d'avantage.

Mais elle ne peut posséder un nombre infini de connaissances, ni en acte, ni à l'état habituel.²⁵

L'infini *en puissance* est un infini par succession : nous pouvons toujours diriger notre attention vers de nouveaux objets que nous n'avions pas considérés avant. Appliquons cela aux désirs : nous pouvons diriger notre attention vers un désir que nous avons eu, puis le juger bon ou mauvais. Nous pouvons même porter notre attention sur un désir que nous avons encore pour l'évaluer, ce qui permet de désirer ce désir (ou de le vouloir). Mais la question est de savoir si nous pouvons avoir un désir de niveau 5, 6, ou d'un niveau plus élevé encore. Un tel désir suppose de viser consciemment le désir de niveau immédiatement inférieur. Ce qui nous intéresse est la connaissance *actuelle* ou *habituelle* de ce désir, et non une succession de connaissances plus ou moins juxtaposées.

Pour désirer-6 une barre de chocolat, il faut se représenter d'une manière ou d'une autre l'objet du désir : le désir 5 de la barre de chocolat (le désir de niveau 6 ne porte pas sur la barre de chocolat, puisque seul un désir de niveau 1 peut porter sur autre chose qu'un désir).

²⁵ Thomas d'Aquin, *Somme théologique*, I, Q.86, a.2, p. 754.

Or il est psychologiquement improbable que nous puissions faire cela. Nous pouvons sans doute dessiner un schéma pour nous aider à distinguer le désir 5 et le désir 6 correspondant, mais cette représentation purement abstraite ne nous aiderait nullement à imaginer cette différence à travers des situations concrètes où il y aurait une opposition entre les deux niveaux. Ce serait une représentation purement conceptuelle. Ajoutons que pour se représenter le désir 6, il faut se représenter en même temps tous les désirs correspondants des niveaux inférieurs, puisque les désirs des niveaux supérieurs n'existent que par rapport à ceux des niveaux inférieurs. Or comment pourrions-nous viser simultanément des objets à la fois si nombreux et si abstraits ?

Reste la connaissance *par habitus* : pouvons-nous avoir des désirs de niveau supérieur à 2 ou 3 au moyen d'une connaissance habituelle des niveaux inférieurs ? La connaissance *par habitus* suppose préalablement une connaissance *en acte*. Par exemple, je connais de manière *habituelle* la configuration des bâtiments de l'université : je sais que le RU se trouve à l'ouest de la faculté de lettres, à côté de la ligne de tramway, etc. Quand je vais à la faculté de lettres en sortant du tramway, je n'ai pas besoin de me représenter *actuellement* qu'elle se trouve à l'est. J'y vais, tout simplement, ma connaissance n'ayant pas besoin d'être une connaissance actuelle pour être efficace (il suffit qu'elle soit mémorisée). Ici, nous voyons que la connaissance habituelle est le résultat d'une connaissance actuelle mémorisée. C'est parce que j'ai su de manière *actuelle* que la faculté de lettres était placée ainsi par rapport à l'arrêt de tramway que je le sais ensuite de manière *habituelle*. De même, pour avoir une connaissance habituelle d'un désir de niveau 5, il faut en avoir eu d'abord une connaissance actuelle. Ainsi, nous n'avons pas fait un pas de plus dans la résolution de la difficulté : les désirs de niveaux supérieurs à 2 ou 3 supposent une forme de conscience apparemment étrange et inaccessible : la connaissance actuelle de tous les désirs correspondants dans les niveaux inférieurs, et de l'objet du désir 1, par rapport auquel tous les niveaux supérieurs prennent leur sens. Nous imaginons facilement des situations où il y a un conflit ou une convergence entre les désirs des deux premiers niveaux. Par exemple, je désire résister à mon désir de manger une barre de chocolat, ou au contraire, je désire céder à ce désir, voire en être complètement esclave, afin d'avoir une bonne excuse pour le satisfaire. Mais dès que nous essayons de passer à un niveau supérieur pour imaginer des conflits entre les niveaux 2 et 3, puis 3 et 4, nous constatons qu'il est difficile de trouver des exemples. Nous voyons bien qu'il y a en principe une différence

conceptuelle entre un désir de niveau 3 et un désir de niveau 4, mais nous ne voyons pas *concrètement* quelle différence il y a entre ces désirs. Eleonore Stump a imaginé un conflit entre les niveaux 2 et 3, mais c'est un cas limite, et il ne semble pas qu'on puisse monter beaucoup plus haut :

Considérons quelqu'un qui a gardé depuis son enfance chez les baptistes, dans le Sud, un désir d'éviter l'alcool, et un désir de second ordre d'avoir ce désir. Imaginons qu'ensuite, cet individu renie la religion de son enfance et, entre autres choses, rejoigne un club de voile où on sert régulièrement de la bière après les courses. Il aura donc un désir de premier ordre de boire de la bière pour s'adapter, être de compagnie agréable, et peut-être aussi céder à une récente attirance pour la bière ; mais il peut aussi, par un reste des inclinations de son enfance, trouver en lui-même un sentiment de culpabilité lorsqu'il boit, et un désir de second ordre d'avoir cette sorte de volonté par laquelle on ne veut pas boire. Ensuite, il se rappellera qu'il a renié son passé baptiste et qu'il n'y a pas de mal, mais plutôt même un bien positif, dans le fait de boire quelques bières. Dans ce cas, donc, il aura un désir de troisième ordre de ne pas avoir le désir de second ordre légué par son éducation²⁶.

Comme l'a écrit Stump, les conflits de ce genre ne durent pas longtemps. Il s'agit de toute façon d'un cas limite en dehors duquel les désirs de niveau 3 se confondent avec ceux du niveau 2. Ce cas limite suppose que le désir de second ordre échappe plus ou moins au contrôle conscient exercé par l'agent sur ses propres désirs. Entre les désirs sensibles de type animal et les volontés clairement conscientes, il y a une place pour des inclinations relevant à la fois de la sensibilité et de l'intellect. Il y a donc deux sortes de désirs, plus une catégorie intermédiaire. Les désirs purement sensibles ne font pas nécessairement l'objet d'un désir de niveau 2. Dans le *De Anima*, Aristote admet l'existence d'un sentiment de sentir chez les animaux : en sentant un objet brûlant, par exemple, l'animal sent qu'il sent²⁷. Mais cette sensation de second ordre ne suppose pas un objet supplémentaire que l'animal viserait distinctement. Chez l'homme, il y a, au-dessus des désirs purement sensibles, d'autres désirs impliquant l'intelligence et la volonté, mais bien qu'il soit doué de raison, l'homme ne vit et ne pense pas de façon parfaitement rationnelle. Chez lui, les désirs des niveaux supérieurs subissent donc l'influence des désirs sensibles, et il se peut que des désirs de niveau 2 ou 3

²⁶ Eleonore Stump, « Intellect, Will, and Alternate Possibilities » (1990), in Fischer et Ravizza (1993), p. 221.

²⁷ Aristote, *De l'Âme*, III, 2, 425b 12 et sq. (p. 204 dans la traduction française par Richard Bodéüs, GF-Flammarion). Dans ce passage, Aristote vise probablement à la fois l'homme et les animaux supérieurs.

aient encore quelque chose de sensible et d'irrationnel. D'où la possibilité de conflits à un niveau relativement élevé, entre des désirs déjà à moitié rationnels et d'autres qui le sont d'avantage. Rien n'empêche donc qu'il y ait des conflits à des niveaux très élevés, si ce n'est une certaine faiblesse de l'esprit de l'homme et de sa conscience.

Lorsqu'il y a une opposition entre le niveau 5 et le niveau 6 (dans l'hypothèse où l'agent est doté d'une conscience suffisamment puissante pour rencontrer ce genre de problèmes), le désir 5 et le désir 6 constituent évidemment deux désirs différents. Au contraire, lorsqu'il n'y a aucune tension entre les désirs des niveaux successifs, les désirs des niveaux supérieurs se confondent avec ceux des niveaux inférieurs ; il n'est donc pas nécessaire d'imaginer que nous visions une infinité d'objets distincts. Au fond, pour qu'un agent ait un désir 6 de manger une barre de chocolat, il suffit qu'il désire consciemment une barre de chocolat. A partir du moment où son désir est conscient de lui-même et s'approuve lui-même (donc tant qu'il n'y a pas de conflit d'inclination), il se redouble indéfiniment. Ainsi, au lieu de représenter cela comme un empilement infini de désirs orientés vers le même objet, il faut se rappeler que le désir de niveau 2 est réflexif ou conscient de lui-même, c'est pourquoi il n'y a en réalité qu'un seul désir mais qui s'approuve lui-même.

Lorsqu'il y a conflit entre deux désirs de niveaux différents, comme dans l'exemple de Stump, il n'y a plus un seul objet de conscience, mais deux, trois, ou d'avantage, car désirer une bière et désirer résister à ce désir sont deux actes fort différents, qui ne portent pas sur le même objet. On ne peut pas dire qu'un de ces actes soit contenu dans l'autre. La faiblesse de notre conscience, incapable de se donner pour objet un désir de niveau 5 ou 6, limite ainsi la complexité de notre psychologie. Inversement, on peut supposer qu'un individu doté d'une conscience plus puissante maîtriserait beaucoup mieux ses désirs, ce qui lui permettrait de les accorder, de telle sorte qu'il serait moins tiraillé entre des désirs contraires.

En tout cas, un fait mérite notre attention : le désir sensible n'est pas réflexif comme peut l'être la volonté. La volonté est un appétit dirigé par une connaissance conceptuelle. Elle est donc liée à l'intellect, qui nous représente certaines choses comme bonnes et d'autres

comme mauvaises. Or il semble que notre volonté, lorsqu'elle agit de façon consciente²⁸ (comme dans un choix ou une décision), soit profondément réflexive, de telle sorte qu'en choisissant quelque chose, nous choisissons *ipso facto* de faire ce choix, et ainsi de suite. Contrairement au *désir* de manger du chocolat, le *choix* d'en manger est un acte essentiellement conscient de lui-même. On peut *désirer* fumer sans en être conscient, mais on ne peut pas *choisir* de fumer sans en être conscient. C'est pourquoi il est possible d'avoir des désirs de premier niveau sans qu'ils soient redoublés à un niveau supérieur, alors que tout choix est nécessairement redoublé à l'infini : choisir *x*, c'est en même temps vouloir choisir *x*, vouloir vouloir choisir *x*, etc. Les désirs sensibles sont liés à la perception, qui n'est pas nécessairement consciente, parce qu'elle a sa cause en dehors de l'agent. Au contraire, l'agent est la cause première de ses choix, c'est pourquoi il ne peut les avoir sans être conscient qu'il les a. C'est ce qu'on appelle une connaissance *pratique*.

Il y a des choses que nous ne *savons* que parce que nous les *faisons*. Cette connaissance est alors une connaissance *pratique*²⁹. Par exemple, si j'écrivais une nouvelle, ce n'est pas en me voyant utiliser le clavier de mon ordinateur que je le saurais. Je sais que j'écris une nouvelle parce que je fais cela *librement et intentionnellement*. Imaginons maintenant que j'aie le hoquet. C'est en m'observant moi-même que j'apprends cela, et non en agissant, car le fait de hoqueter n'est pas une action mais un comportement involontaire. Cette deuxième connaissance n'est donc pas une connaissance pratique.

De même, lorsqu'un désir naît indépendamment de notre volonté, c'est une tendance dont nous n'avons pas conscience immédiatement, mais que nous constatons après qu'elle soit née, alors que nos désirs volontaires (ou plus simplement nos volontés) sont des inclinations dont nous sommes la cause première. Elles sont des sortes d'actions, et comme toute action,

²⁸ Il est possible qu'un comportement volontaire devienne habituel et cesse par là d'être conscient. On peut oublier ce qu'on veut faire tout en continuant à le faire. Peut-être même est-ce possible, en dehors du cas des *habitus*, de faire une action sans savoir pourquoi on la fait, c'est-à-dire sans savoir ce qu'on veut en la faisant. Par exemple, une jeune fille réveille son frère en entrant dans sa chambre sous prétexte qu'elle a besoin de comprimés d'aspirine, mais contrairement à ce qu'elle pense, elle veut surtout réveiller son frère et lui parler.

²⁹ Sur la connaissance pratique, voir Vincent Descombes, « Comment savoir ce que je fais ? », in *Philosophie* n°76 (décembre 2002). Descombes fait lui-même référence à des réflexions de Thomas d'Aquin, Wittgenstein, et Elizabeth Anscombe – cf. Elizabeth Anscombe, *L'Intention* (1957), tr. M. Maurice et C. Michon, Paris, Gallimard, 2002.

nous les connaissons par connaissance pratique, c'est-à-dire par le simple fait que nous en sommes l'origine.

En résumé, lorsqu'un désir est dépourvu de conscience, il n'est pas contrôlé par l'agent et ne fait pas l'objet d'un désir supérieur ; et lorsqu'il est clairement conscient, il se prend lui-même pour objet, ce qui donne une sorte de régression non-vicieuse à l'infini. Ainsi, il y a une pluralité de niveaux dans les volitions, mais il faut distinguer deux cas. Lorsque les désirs de niveaux différents ne se contredisent pas, il y a en un sens des volitions de niveaux 3, 4, 5, etc., mais comme le dit Stump, les désirs situés au-dessus du niveau 2 n'ont pas d'existence propre (ou d'objet propre), ils ne sont au fond qu'un seul désir conscient de lui-même et qui s'approuve lui-même. Au contraire, lorsque les désirs des différents niveaux sont en désaccord, il n'y a plus *un seul* objet de désir, mais *plusieurs*, ce qui exige une capacité d'attention plus élevée.

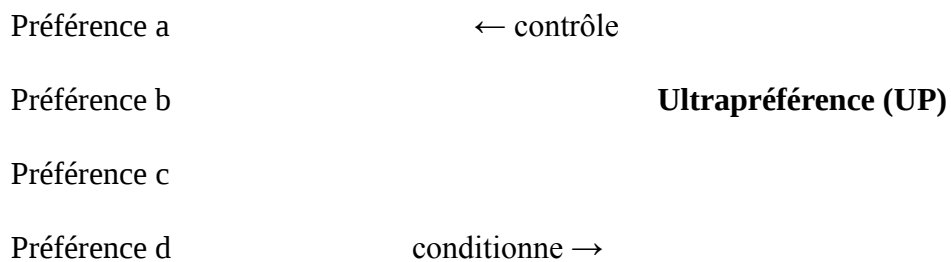
4.5 Application au compatibilisme

Voyons maintenant si une théorie compatibiliste peut définir la liberté et la responsabilité morale en tenant compte de ces données. Keith Lehrer propose une théorie compatibiliste qui résout le problème de la régression à l'infini à peu près comme je viens de le faire, c'est-à-dire en concevant la régression à l'infini sous la forme d'un cercle. D'après lui, « l'ajout de préférences au-delà du second ou du troisième niveau, bien qu'elle puisse être opérée sur un mode calculatoire (*computationally*), manquerait d'efficacité causale »³⁰. Par exemple, le raisonnement (le calcul) nous fait dire que celui qui choisit d'arrêter de fumer a du même coup la volonté de vouloir vouloir arrêter de fumer ; mais nous avons le sentiment que cela n'ajoute rien, du point de vue causal, à son choix d'arrêter de fumer. Lehrer admet donc une infinité de niveaux, mais à condition de les considérer comme différents seulement du point de vue conceptuel, et non du point de vue causal. Tous ces désirs ensemble forment une structure de désirs (ou « structure de préférences ») liée à des croyances qui justifient ces

³⁰ Keith Lehrer, « Freedom and the Power of Preference », in *The Determinism and Freedom Website*, ed. Ted Honderich. Cette phrase se trouve p.4 de l'article de Lehrer.

désirs ; c'est pourquoi nous pouvons l'appeler aussi « structure de raisons ». Or l'agent est capable de vouloir ou d'approuver cette structure toute entière en un même acte de volonté. C'est ce que Lehrer appelle « l'ultrapréférence ». Si l'agent n'est pas content de sa structure de préférences, il peut s'efforcer de la rendre plus cohérente en luttant contre ses penchants qui lui paraissent mauvais. La structure de préférences dépend donc de l'ultrapréférence³¹. Or celle-ci dépend en retour de la structure de préférences. Il y a une circularité causale entre les deux :

Structure de préférences (SP)



Comment la structure de préférences conditionne-t-elle nos préférences ? Chaque désir ou préférence dépend des désirs des premiers niveaux. Par exemple, si je choisis de céder à mon désir de fumer une cigarette, c'est parce que j'ai un désir de premier niveau qui me porte à fumer. De même, l'ultrapréférence correspond à un désir appartenant à la structure de préférences (un désir sensible, une valeur morale, ou autre chose encore). Il y a des choix qu'on ne peut pas faire, parce qu'on n'a pas les désirs correspondants : on ne peut pas désirer succomber à une tentation si on n'a pas cette tentation, et on ne peut pas désirer la vaincre si on n'a aucune valeur ou tendance qui s'y oppose.

Cela dit, la relation entre les désirs et l'ultrapréférence est obscure à cause de sa circularité. Il ne s'agit pas d'une relation d'implication logique, mais d'une relation *causale*. Or comment l'ultrapréférence pourrait-elle causer la structure SP si celle-ci est la cause de celle-là ? La causalité peut-elle tourner en cercle, alors que la cause est normalement antérieure à l'effet ?

³¹ *Ibid.*, p. 9.

Que répondre à cela ? Pas besoin de croire à un Éternel Retour. En fait, nous connaissons de nombreux cas de causalité réciproque. Lehrer donne l'exemple d'une arche ³²: dans cet assemblage, la clé de voûte est maintenue en hauteur par les pierres des arches appuyant de part et d'autre, et celles-ci ne tiendraient pas si l'on retirait la clé de voûte. Ainsi, il y a dépendance réciproque entre l'arche et la clé de voûte. L'arche représente le système de préférences, et la clé de voûte, l'ultrapréférence. C'est là une métaphore, bien sûr, mais qui prouve néanmoins l'existence de la dépendance causale réciproque.

La solution tient donc dans cette idée : la structure de préférences (SP) dépend de l'ultrapréférence (UP) qui dépend de SP, qui dépend de l'ultrapréférence, et ainsi de suite. Ce « cercle vertueux » donne l'impression que les préférences de l'agent n'ont pas de cause extérieure à lui. Il les contrôle lui-même, c'est pourquoi il ne doit qu'à lui-même ³³ d'avoir des préférences qui le font agir de telle ou telle façon.

En réalité, cette impression est fallacieuse, et d'ailleurs, Lehrer lui-même ne s'y arrête pas ³⁴. Si je fabrique un appareil ou un système comprenant des relations d'interdépendance, le fait que plusieurs éléments du système soient interdépendants n'exclut pas qu'ils soient entièrement déterminés par des causes extérieures. Par exemple, le thermostat d'un frigidaire permet de maintenir sa température à un niveau constant parce qu'il réagit en fonction des variations de température à l'intérieur du frigidaire. Le comportement du thermostat dépend de la température ambiante, qui dépend du thermostat, et ainsi de suite. Pourtant, nul ne doute que ces choses sont des assemblages de matière inerte absolument dépourvus d'autonomie et de liberté. D'ailleurs, le monde vivant est structuré par une multitude d'interactions sans que la liberté et la responsabilité y soient omniprésentes. Les animaux transforment leur milieu naturel, mais s'y adaptent en fonction des informations qu'ils reçoivent du monde extérieur

³² *Ibid.*, p. 9.

³³ Lorsqu'il est responsable moralement.

³⁴ cf. *ibid.*, p. 11 : Lehrer se demande ce qui vient en premier : la structure de préférences ou l'ultrapréférence ? Il fait l'hypothèse que c'est l'esprit qui se forme d'abord, puis le « méta-esprit ». Un être humain possède donc des désirs et des préférences avant d'avoir une ultrapréférence. Et ces désirs sont déterminés causalement. Pour cette raison, Lehrer est amené à répondre à l'argument du transfert quelques pages plus loin (p.14-15). Pour cela, il renvoie à l'article de D. Lewis dont j'ai résumé l'argument dans le chapitre III (section 2) : David Lewis, « Are we free to break the laws ? » (1981).

grâce à leurs sens, et leurs réactions sont infiniment plus variées que celles d'un thermostat. A l'intérieur même d'un animal, les différents organes sont interdépendants : les poumons et le système digestif dépendent du cœur qui leur envoie du sang, mais le système circulatoire dépend à son tour de la composition du sang, donc des poumons et du tube digestif. La notion d'interaction est trop vague pour définir la liberté, aussi n'est-elle pas le principe de la solution. L'idée essentielle, chez Frankfurt et Lehrer, est bien que la liberté consiste à vouloir librement et à maîtriser ses désirs.

Résumons ce qui précède. Le compatibilisme s'appuie souvent sur une définition conditionnelle de la liberté. Or cette définition n'est pas suffisante parce qu'elle s'applique seulement aux actions et non à la volonté ou au désir. Pour la compléter, il faut donc la redoubler au niveau des désirs de l'agent. Mais cela implique des désirs de second ordre, et ceux-ci, à leur tour, ne paraissent libres que s'il y a des désirs de 3^{ème} niveau dont ils sont l'objet. Cela nous entraîne dans une régression à l'infini difficilement concevable, à moins de supposer que les choix volontaires puissent être conscients d'eux-mêmes. Ainsi, la régression à l'infini prend la forme d'un cercle et elle s'explique par la réflexivité d'une certaine forme de conscience.

4.6 Contrôle de soi et identité avec soi

Il y a des cas où, de toute évidence, l'agent n'est pas responsable de ses propres désirs. Par exemple, un individu qui a été drogué contre sa volonté peut devenir dépendant. Quand il est en manque, on ne peut pas lui reprocher de consommer de la drogue, car cette action est l'effet d'un désir qui, au fond, ne vient pas de lui, mais d'une influence extérieure. La théorie du « moi profond » permet de rendre compte de tous les cas de ce genre, où l'individu, par exemple, agit sous l'effet d'une drogue, de l'hypnose, d'un lavage de cerveau ou d'une maladie mentale. On appelle « théorie du moi profond » une théorie compatibiliste selon laquelle la responsabilité de l'agent est liée au fait que ses désirs et ses actions expriment son « moi profond ». Frankfurt, Watson, et Taylor partagent l'idée que si notre caractère découle simplement de l'hérédité, de l'influence du milieu, ou de causes extérieures échappant à notre contrôle, nous ne pouvons pas être considérés comme responsables de nos actions. De même,

les animaux ou les enfants en bas âge n'ont pas de responsabilité morale dans ce qu'ils font, parce qu'ils ne sont pas responsables de leurs inclinations.

Il faudrait donc être responsable de son moi pour être responsable de ses actes. Contrairement à Watson et Frankfurt, Taylor³⁵ ne s'appuie pas sur une distinction entre plusieurs niveaux d'inclinations. Il admet simplement que l'agent doit avoir la capacité de réviser ou de corriger ses propres désirs. En un sens, cette idée rejoint la même intuition : l'agent est responsable quand il est cause de lui-même. A partir de là, on aboutit naturellement à l'idée que l'agent a des désirs ou des volitions de second niveau, par lesquelles il veut avoir ou ne pas avoir tels ou tels désirs. Les désirs du premier niveau correspondent au moi superficiel de l'agent, du moins s'il ne les approuve pas. Et les désirs de second niveau manifestent la volonté profonde de l'agent, c'est-à-dire son moi profond (ou sa *personnalité*).

L'objection du transfert, que j'ai défendue au chapitre III, consiste à dire que si tout est déterminé par des causes antérieures à notre naissance, puisqu'elles échappent évidemment à notre contrôle, tout ce qui en découle nous échappe, même nos propres actions. Selon Suzan Wolf³⁶, « la théorie du moi profond offre une réponse » à cette objection, « car elle nous autorise à distinguer les cas où les désirs sont déterminés par des forces étrangères au moi, et ceux où ils sont déterminés par le moi lui-même, c'est-à-dire par des désirs de second ordre, par le moi « réel » ou « profond » ». Mais « c'est une réponse qui laissera beaucoup de gens insatisfaits. Même si mes actions sont gouvernées par mes désirs et mes désirs par mon moi plus profond, il reste une question : qui est responsable de ce moi plus profond ? Apparemment, la réponse donnée plus haut ne fait que repousser le problème »³⁷.

Pour répondre à cette objection, on peut soutenir que l'agent est aussi la cause de son moi profond. D'un point de vue déterministe, cette idée peut prendre deux formes différentes. Tout d'abord, il est clair qu'on ne peut remonter indéfiniment dans le passé d'un individu,

³⁵ Charles Taylor, « Responsibility for Self », in A. E. Rorty, ed. *The Identities of Persons*, University of California Press, Berkeley, 1976.

³⁶ Suzan Wolf, « Sanity and the Metaphysics of Responsibility », from *Responsibility, Character and the Emotions* (1987), p.42-62, Cambridge University Press, reproduit dans *Free Will*, anthologie d'articles sur le libre arbitre réunis par Gary Watson, Oxford University Press, Oxford, 2003.

³⁷ *Ibid*, p. 377.

pour montrer que chacun de ses désirs ou de ses actes est l'effet d'un désir ou d'un acte antérieur. Pour cela, il faudrait que l'individu ne soit jamais né, qu'il soit sans commencement. L'autre solution consiste à imaginer une infinité de niveaux dans les inclinations. On suppose alors que la liberté de l'agent est liée à sa capacité d'approuver ou de vouloir ses propres désirs : ses désirs sont volontaires, mais aussi son désir de désirer telle chose, et ainsi de suite - je ne reviens pas sur cette difficulté.

Avec Lehrer, nous avons vu comment un compatibiliste pouvait résoudre le problème de la régression à l'infini : en la réduisant à un cercle. Chez Lehrer, ce cercle est dans la dépendance réciproque entre les préférences et l'ultrapréférence. Mais il faut bien que l'un de ces deux éléments apparaisse en premier. Lehrer suppose que l'esprit apparaît avant le « méta-esprit » ; autrement dit, tout être humain a d'abord des *désirs*, puis des *préférences au sujet de* ses désirs. Du coup, on voit qu'il n'est pas la cause ultime de ses désirs. Ils ne viennent pas de ses préférences, puisqu'elles n'existent pas encore, mais l'agent naît avec eux sans les avoir choisis. C'est ici que la théorie de Lehrer rencontre inévitablement l'objection du transfert. Il y répond de la même manière que Lewis, mais nous avons vu (dans le chapitre III) que cette réponse n'était pas satisfaisante. Les arguments de Lewis, Slote, et Fischer ne réfutent pas l'argument du transfert si celui-ci est bien formulé.

Dans le chapitre III, j'ai retenu les formulations suivantes du transfert d'impuissance :

(TCI 2) $N^W p, p \rightarrow q$, donc $N^W q$. (O'Connor)

(TCI 3) $N^W p, N^W(p \rightarrow q)$, donc $N^W q$. (McKay et Johnson)

Elles ne peuvent être dérivées validement du transfert de capacité correspondant, mais j'ai proposé une autre formulation qui se dérive du principe suivant :

(TCC 5) $A^C p, (p \text{ et si } p, p \text{ cause } q)$, donc $A^C q$.

(A est causalement capable de faire que p, (p et si p, p cause q), donc A est causalement capable de faire que q).

De là, on peut déduire :

(2) $A^{C\sim}q$, (p et si p, p cause q), donc $A^{C\sim}p$.

(3) $\sim A^{C\sim}p$, (p et si p, p cause q), donc $\sim A^{C\sim}q$.

(TCI 5) $N^c p$, (p et si p, p cause q), donc $N^c q$.

Ces trois versions de TCI permettent d'expliquer le principe du transfert de non-responsabilité :

(TCNR) Si $\sim R p$, et $\sim R(p \text{ cause } q)$, alors $\sim R q$.

Appliquons cela à un agent quelconque S :

Si $\sim R^s p$, et $\sim R^s(p \text{ cause } q)$, alors $\sim R^s q$.

Si on estime que pour être responsable d'un événement, il faut en être la cause (donc si on pense que $R^s q \rightarrow Csq$), il en découle qu'il faut être la cause ultime de cet événement. Sinon, on remontera à des causes de plus en plus éloignées dans le passé, jusqu'à des causes antérieures à la naissance de l'agent. De même, si l'agent n'est responsable de ses désirs que lorsqu'ils résultent de ses désirs précédents, il est clair qu'en dernière analyse, il n'est responsable d'aucun de ses désirs, car il est né avec des désirs dont il ne pouvait lui-même être la cause. Puisque la chaîne des désirs a eu un commencement, et puisque les premiers désirs de l'agent ne sont pas l'effet de ses propres désirs, il ne peut être considéré comme vraiment autonome.

Les théories de Watson et de Lehrer présupposent au fond une même idée du contrôle de soi-même. Pour Lehrer, il faut que les désirs de l'agent s'identifient à son ultrapréférence, que nous pourrions définir comme le cœur de son identité. Ce qui importe, d'après Watson, c'est que l'agent s'identifie avec ses propres désirs ; et cela arrive si ses désirs correspondent à ses valeurs. L'agent devient pleinement libre lorsque sa structure motivationnelle coïncide avec sa structure de valeurs. Nous voyons donc que pour Watson, la structure de valeurs est ce qui constitue le moi profond de l'agent, le noyau de son identité personnelle. Un tel

présupposé se heurte naturellement à une objection : pourquoi penser que les valeurs de l'agent constituent le fond de son identité personnelle ? Après tout, les valeurs ont souvent un caractère impersonnel : elles sont transmises par l'éducation, partagées par un milieu social ou même par une société entière. D'ailleurs, nous avons tendance à penser que les valeurs les meilleures sont les plus universelles, donc les moins personnelles. Même si le moi profond consiste en un désir singulier, propre à un seul individu, le fait qu'il soit produit par des causes indépendantes de l'agent pose un sérieux problème.

Que peut être le moi de l'agent ? Il est difficile de définir l'identité personnelle – j'entends par là, non pas l'identité à travers le temps, par exemple entre Marc à t1 et Marc à t2, mais ce qui définit en quelque sorte cet individu à n'importe quel moment³⁸. Ce que nous devons chercher ici, ce n'est pas si l'agent est une partie de lui-même (un « homoncule »), mais où est le centre de son identité personnelle. Je n'exclus d'ailleurs pas qu'il y ait plusieurs sphères de l'identité personnelle à l'intérieur de l'agent ; ce que je veux dire, c'est que les propriétés et les parties qui nous constituent n'ont pas toutes la même relation à notre identité profonde. Par exemple, ma personnalité, qui constitue en un sens mon identité, comprend des traits de caractère et des tendances que je ne contrôle pas entièrement. On n'est pas responsable du fait qu'on soit homme ou femme, noir ou blanc, émotif ou non, or tout cela fait partie dans une certaine mesure de l'identité des personnes. Mais il existe en nous un

³⁸ Notons que ces deux formes d'identités sont liées. Par exemple, nous supposons que l'orteil de Marc ne fait pas essentiellement partie de lui parce que s'il le perd, il reste la même personne. Plusieurs théories sur l'identité personnelle à travers le changement proposent comme critère la permanence d'une certaine partie de la personne à travers le temps : son âme, son cerveau, ou une partie de son cerveau. Swinburne suppose qu'il s'agit de son âme ; Van Inwagen pense qu'il s'agit plutôt d'une partie du cerveau, mais non le cerveau entier, car un individu ayant perdu une partie de son cerveau au cours d'une opération est toujours considéré comme le même individu. Pour Locke, le critère de l'identité personnelle est phénoménal : c'est la continuité dans les états de conscience. Si un individu perd complètement la mémoire, il cesse pour ainsi dire d'exister, même si son corps est toujours en vie, il y a donc deux personnes successives dans le même corps (Locke, *Essai sur l'Entendement humain* II, XXVII). Leibniz et Richard Swinburne pensent que la continuité psychique ne constitue pas l'identité personnelle, même si elle en est un critère. Elle suppose la permanence d'une partie de l'individu : son âme, et c'est là que réside l'identité personnelle (cf Leibniz, *Nouveaux Essais*, II, ch.XXVII, et Sidney Shoemaker & Richard Swinburne, *Personal Identity*, p.65). Van Inwagen préfère s'en tenir au matérialisme et soutient que l'identité personnelle à travers le changement suppose la permanence d'une certaine partie du cerveau. Les deux hémisphères cérébraux n'ont pas les mêmes fonctions ni la même importance. Il est donc possible de croire que l'identité personnelle est liée à une partie du cerveau seulement. Van Inwagen résout ainsi les difficultés qui se posent dans l'expérience de pensée des hémisphères cérébraux dissociés, ou face aux cas de personnes qui se font enlever une partie du cerveau atteinte d'une tumeur. Après l'opération, nous pensons avoir affaire aux mêmes personnes, même si leurs facultés sont diminuées. C'est peut-être parce que l'opération n'a pas altéré ce qui est essentiel, dans le cerveau, pour l'identité personnelle (Van Inwagen, *Material Beings*, section 16, p.190).

noyau de notre identité qui est lié à notre libre arbitre. En effet, ce que chacun de nous contrôle en lui-même (et dont il est pleinement responsable), est également ce qui exprime le mieux *ce qu'il est*. Le principe suivant paraît donc solide :

(IP): Ce qui constitue l'identité profonde de A (le noyau de son identité personnelle), c'est ce que A contrôle immédiatement.

Un lien étroit relie donc notre liberté à une certaine partie de nous-mêmes, noyau ou point de départ de notre liberté. Ce noyau est constitué par ce qui fait partie de nous et que nous contrôlons le plus directement. Il est ce sans quoi nous ne pouvons rien contrôler du tout, ni en nous-mêmes, ni dans le monde extérieur. Nous pouvons perdre nos mains sans cesser de contrôler une partie de nous-mêmes et du monde extérieur, mais sans notre cerveau, par exemple, que pourrions-nous faire ? Pour être, il faut agir, et donc contrôler quelque chose. Ce noyau, qu'il s'agisse ou non du cerveau, est donc en même temps ce que nous ne pouvons pas perdre sans cesser d'exister. Tout ce qu'un agent contrôle de façon indirecte à partir de son cerveau, il peut le perdre sans cesser d'exister. Mais en admettant qu'il contrôle directement son cerveau³⁹, il est inconcevable que celui-ci soit détruit sans que l'agent soit également détruit. Bien entendu, cela ne veut pas dire que l'agent se réduise à son cerveau, mais que celui-ci est le centre de son identité personnelle : ce qui ne peut disparaître sans que l'agent disparaisse également.

Il semble que toutes les théories déterministes ou compatibilistes soient amenées à supposer un critère du même genre que celui de Watson. Or s'il paraît évident que l'agent s'identifie avec ce qui répond immédiatement à sa volonté (puis dans une moindre mesure avec les autres parties de son être), il n'est pas du tout évident, au contraire, que l'agent s'identifie avec ses valeurs, ses croyances, ou sa conscience, ou bien encore son inconscient. IP est un critère minimal exprimant une forte intuition à laquelle il serait difficile de renoncer, alors que le critère admis par Watson a quelque chose d'arbitraire. Sur ce point, les thèses libertariennes présentent un avantage : elles peuvent se contenter de IP comme critère de

³⁹ Ou qu'il ne peut rien contrôler sans son cerveau.

l'identité personnelle (entendons par là : l'identité personnelle dont il s'agit quand nous parlons de l'autonomie et de la responsabilité morale).

Leur deuxième avantage est qu'elles s'accordent avec le principe du transfert de non-responsabilité. Que le moi profond de l'agent se constitue de désirs, de croyances, de valeurs, ou d'autre chose encore, si cela découle entièrement d'événements passés en vertu des lois de la nature, l'agent ne le contrôle pas et n'en est pas responsable, ni des désirs et des actions qui en découlent.

Résumé. Dans les sections précédentes, j'ai examiné deux stratégies compatibilistes. La première et la plus classique est de formuler une définition conditionnelle de la liberté pour montrer qu'elle est parfaitement compatible avec le déterminisme causal. Mais comme nous l'avons vu, la responsabilité morale exige une capacité de contrôler ses propres désirs, et cela conduit à une deuxième stratégie (complémentaire de la première) selon laquelle la liberté de l'agent suppose qu'il puisse vouloir ses propres désirs. Or pour que ce vouloir de second ordre soit lui-même libre, il faudrait continuer ainsi à l'infini, ce qui est difficilement concevable. On évite cette difficulté en montrant que le vouloir peut être réflexif : un *choix* est conscient de lui-même, donc se prend lui-même pour objet. Mais il reste une objection à laquelle ni Lehrer, ni les autres, ne semblent répondre de façon satisfaisante : c'est l'objection du transfert, que j'ai défendue dans le troisième chapitre. La théorie hiérarchique des désirs veut montrer que l'agent peut avoir la responsabilité de ses désirs profonds, mais si le déterminisme est vrai, il est impossible que nous ayons la responsabilité *ultime* de nos désirs.

Je vais maintenant aborder une troisième stratégie compatibiliste, qui se fonde plutôt sur l'idée que la responsabilité morale est liée à la *rationalité* de l'agent. Wolf et Fischer renoncent d'emblée à l'idée d'une responsabilité ultime et préfèrent chercher le critère de la responsabilité dans l'aptitude de l'agent à se comporter de façon rationnelle (c'est-à-dire morale).

4.7 Liberté et rationalité

Au début du chapitre I, nous vu qu'il y avait principalement trois manières de concevoir la liberté et le libre arbitre : être libre, c'est faire ce qu'on veut, agir par soi-même, ou avoir la possibilité d'agir autrement. Pour examiner les théories compatibilistes, nous sommes partis de la première idée, qui nous a conduit à la deuxième, mais nous avons vu que celle-ci était également problématique. Les théories qui vont nous intéresser maintenant sont fondées plutôt sur la troisième idée : la responsabilité de l'agent suppose une capacité d'agir autrement. Nous partirons de Wolf, mais je discuterai surtout la théorie de Fischer et Ravizza, plus élaborée.

D'après Suzan Wolf, le déterminisme exclut l'idée d'autonomie ultime⁴⁰, car si le déterminisme est vrai, aucun agent humain n'est une cause première de ses propres états ou inclinations. Même en les expliquant à partir de causes internes à l'agent (désirs, croyances, etc.) il est nécessaire d'admettre, en dernière analyse, que tout cela est le résultat de causes extérieures à lui. Il faut donc renoncer au critère de l'autonomie.

Selon Wolf, même si l'autonomie de l'agent est comprise comme un accord de ses actions avec ses désirs profonds, ses valeurs, ou ses désirs de niveau supérieur, elle ne peut constituer un critère suffisant. En effet, on peut agir selon ses désirs profonds (ou selon ses valeurs) sans être responsable de ce qu'on fait. Pour le montrer, Suzan Wolf imagine ce contre-exemple :

Jojo est le fils préféré de Jo premier, le dictateur mauvais et sadique d'un petit pays sous-développé. A cause de l'affection particulière du père pour le garçon, Jojo reçoit une éducation spéciale, il est autorisé à accompagner son père et à observer sa vie quotidienne. Compte tenu de ce traitement, il n'est pas surprenant que le petit Jojo prenne son père comme modèle et développe des valeurs très proches de celles de son papa. Devenu adulte, il fait souvent le même genre de choses que son père : sur un coup de tête, il condamne les gens à mort, à des peines de prison ou à la torture. Il n'est pas contraint de faire cela,

⁴⁰ Wolf parle seulement de l'autonomie, mais pour éviter un risque de confusion, je précise « ultime », comme le fait Kane.

mais agit selon ses propres désirs. De plus, ce sont des désirs qu'il approuve entièrement. Quand il revient en arrière et se demande : « Est-ce que je veux vraiment être ce genre de personne ? », sa réponse est clairement « oui », car ce mode de vie attaché à un fol étalage de puissance constitue son idéal le plus cher⁴¹.

D'après Wolf, Jojo ne peut être considéré comme responsable de ses actes. Il faut au moins lui accorder des circonstances atténuantes, car ses valeurs ont été faussées dès la prime enfance. Le moi profond de Jojo n'est donc pas *sain*, et ce n'est pas à lui qu'il faut attribuer la responsabilité de ce qu'il fait. Or s'il y avait une infinité de niveaux d'inclinations, le problème demeurerait identique. On peut imaginer que Jojo approuve ses désirs de premier niveau, de second niveau, et ainsi de suite, mais cela ne change pas la nature du problème, car les inclinations de Jojo peuvent encore résulter en totalité d'une influence extérieure à lui.

Pour surmonter cette difficulté, Suzan Wolf propose une théorie centrée sur l'aptitude à la rationalité. Selon cette théorie (« *the reason view* » ou « *the sane deep self view* »), le moi profond de l'agent doit être sain ; autrement dit, l'agent doit être *capable d'agir de façon rationnelle*, sans quoi il ne peut être considéré comme responsable de ses actes. La raison est ce qui lui permet de connaître le Vrai et le Bien, puis de se gouverner lui-même en se réglant sur cette connaissance. C'est en cela que réside la véritable autonomie de l'agent ; aussi n'implique-t-elle aucune indétermination, ni même la capacité d'être à l'origine de ses désirs. Si l'agent est autonome, c'est au sens où il est capable d'agir selon la raison, et de se soustraire ainsi à l'influence des passions. Cette idée a une saveur kantienne, et Wolf reconnaît elle-même sa dette à l'égard de Kant⁴², mais contrairement à lui, elle ne pense pas que nous ayons la capacité d'inaugurer des chaînes causales⁴³. Pour Wolf, l'homme est entièrement soumis au déterminisme, et il n'est jamais la cause *première* de quoi que ce soit. L'autonomie de l'agent réside dans sa capacité de se corriger lui-même. La théorie du moi profond et sain reconnaît que pour être responsable de ses actions, il faut être responsable de son moi, ce qui suppose « (1) la capacité de nous évaluer nous-mêmes précisément et

⁴¹ S. Wolf, "Sanity and the Metaphysics of Responsibility" (1987), in G. Watson (2003), p. 379.

⁴² Suzan Wolf, *Freedom within Reason*, Oxford University Press, 1990, p. 10.

⁴³ Emmanuel Kant, *Critique de la Raison pure*, Dialectique transcendantale, troisième conflit des idées transcendantales (p. 348-349 dans la traduction de A. Tremesaygues et B. Pacaud, PUF, Paris, 1965).

judicieusement, et (2) la capacité de nous transformer autant que le demande notre évaluation. Nous pouvons comprendre l'exercice de ces capacités comme un processus par lequel nous *prenons* la responsabilité du moi que nous sommes, mais sans l'avoir créé ultimement »⁴⁴.

En résumé, Jojo serait responsable de ses actes s'il avait la capacité d'agir autrement. Pour cela, il faudrait qu'il ait pu développer son caractère d'une autre manière que dans le monde actuel, et adopter d'autres valeurs que celles qu'il a dans le monde actuel. Or cela exige une raison qui fonctionne correctement.

Comme le reconnaît Suzan Wolf ⁴⁵, sa théorie se caractérise par une asymétrie qui peut sembler curieuse au premier abord. En effet, si un agent n'est pas assez rationnel pour pouvoir agir autrement, il n'est pas responsable de ce qu'il fait, mais si un agent est trop rationnel pour pouvoir agir autrement, par exemple s'il est trop bon pour pouvoir mentir, il est tout de même considéré comme responsable. C'est la thèse de l'asymétrie. Wolf estime qu'une action est méritoire si elle est bonne et si elle a été accomplie pour une bonne raison. Si ces conditions sont satisfaites, l'agent a du mérite même si certaines causes ont déterminé son action. Ainsi, Wolf admet le principe des possibilités alternatives pour les actions mauvaises, mais non pour les actions bonnes.

Pour justifier cette thèse, Wolf présente des exemples de ces deux catégories, en commençant par des cas d'actions louables où l'agent est incapable de faire le mal. Imaginons un garçon qui se débat dans l'eau pour éviter de se noyer, et une femme qui l'aperçoit de la plage. Dès qu'elle l'aperçoit, celle-ci se jette à l'eau pour le sauver. Elle ne peut pas agir autrement car il lui paraît trop évident qu'elle doit agir ainsi. Elle ne peut pas envisager un seul instant de rester passivement sur la plage ; elle est incapable de se représenter une telle possibilité, car son caractère l'en empêche. Selon Wolf, cette femme a du mérite, et est pleinement responsable de ce qu'elle fait. Dans les cas de ce genre, la responsabilité ne

⁴⁴ "Sanity and the Metaphysics of Responsibility", p. 385 dans l'anthologie de Gary Watson.

⁴⁵ Suzan Wolf, *Freedom within Reason*, Oxford University Press, 1990, p. 79.

suppose pas la capacité d'agir de façon moins rationnelle. La responsabilité par rapport à une action bonne ne supposerait donc pas la capacité de ne pas l'accomplir.

Wolf considère ensuite⁴⁶ plusieurs actions mauvaises où l'agent n'a pas la capacité d'éviter le mal. Elle imagine des cas où, sous l'emprise de la folie, d'une drogue ou d'un manipulateur pervers pratiquant l'hypnose, l'agent n'a plus le pouvoir d'éviter le mal. Par exemple, si un kleptomane est submergé par son désir de voler au point qu'il n'a plus la capacité d'y résister, il ne peut être blâmé et jugé responsable de son délit. Il ne suffit pas qu'il ait la capacité de distinguer le bien et le mal. Pour qu'il soit responsable de son action, il faut aussi qu'il puisse agir selon le Vrai et le Bien : « si un agent est incapable de faire les bonnes choses pour les bonnes raisons, alors ce n'est pas sa faute s'il fait un faux pas en commettant le mal »⁴⁷. Wolf conclut qu'il y a une asymétrie entre les bonnes et les mauvaises actions :

La responsabilité dépend de la capacité d'agir selon le Vrai et le Bien. Si quelqu'un est psychologiquement déterminé à faire une bonne chose pour de bonnes raisons, cela n'exclut pas qu'il ait les capacités requises (cela semble en effet être une preuve parfaite qu'il les a). Mais si quelqu'un est psychologiquement déterminé à faire une mauvaise chose, cela semble prouver qu'il ne les a pas, quelles que soient les raisons pour lesquelles il agit.⁴⁸

Cette asymétrie est « paradoxale »⁴⁹, mais selon Wolf, il ne faut pas y voir une contradiction, car nos jugements sur la responsabilité des agents suivent toujours une même règle : il ne sont responsables que s'il ont la capacité de faire la bonne action pour une bonne raison.

⁴⁶ *Ibid.*, p. 83 et s.

⁴⁷ Suzan Wolf, *Freedom within Reason*, p. 81

⁴⁸ Suzan Wolf, *Freedom within Reason*, p. 79.

⁴⁹ *Ibid.*, p. 79.

4.8 Capacités d'agir mieux et d'agir moins bien. Sont-elles asymétriques ?

Fischer et Ravizza⁵⁰ estiment que l'asymétrie mise en avant par Suzan Wolf est une fausse asymétrie. Son erreur viendrait de la manière dont elle choisit ses exemples. Dans ses exemples d'actions irrationnelles (kleptomanie, dépendance, etc.), le choix de l'agent est provoqué par un déterminisme psychologique qui agit *dans le monde actuel*. Or dans les exemples à la Frankfurt, le facteur qui détermine le choix de l'agent n'intervient pas dans la séquence actuelle d'événements, mais dans la séquence alternative. Par exemple, Gunnar tire sur Ridley à *t* sans qu'intervienne le savant fou. Si Gunnar manifestait de l'indécision juste avant *t*, le savant fou déclencherait en Gunnar la décision de tirer sur Ridley. Gunnar tire sur Ridley, qui meurt quelques instants après. Dans cet exemple, l'intervention du savant fou est le facteur qui pourrait exclure la responsabilité morale de Gunnar s'il intervenait dans le monde actuel. Mais puisque ce facteur n'intervient pas dans le monde actuel, Gunnar agit de lui-même. Il est donc responsable de la mort de Ridley. En somme, Wolf aurait dû tenir compte de la spécificité des exemples à la Frankfurt. Pour Fischer et Ravizza, ces exemples prouvent qu'on peut être responsable d'une mauvaise action sans avoir la capacité d'agir autrement ou de ne pas faire cette action. Il n'y aurait donc, selon eux, aucune asymétrie entre les actions bonnes (ou rationnelles) et les actions mauvaises (ou irrationnelles). Dans les deux cas, la responsabilité morale serait liée, non pas à la capacité d'agir autrement, mais au *contrôle de direction (guidance control)*.

Dans le chapitre II, j'ai soutenu que les exemples utilisés pour réfuter les différentes versions du principe des possibilités alternatives⁴ manquent de clarté et même souvent de cohérence. Aucun des ces exemples n'est suffisamment convaincant. Les cas frankfurtiens classiques (avec intervenant contrefactuel) sont réfutés par l'objection du dilemme (formulée par Widerker), et les autres exemples posent également des problèmes. L'objection de Fischer et Ravizza contre la thèse de Wolf repose donc sur une base trop fragile.

⁵⁰ J. M. Fischer & M. Ravizza, *Responsibility and Control*, Cambridge University Press, 1998, p. 55-61.

Si le principe des possibilités alternatives est correct, comme je l'ai soutenu au chapitre II, il est possible de critiquer autrement la thèse de l'asymétrie : en soutenant que la responsabilité morale suppose *dans tous les cas* un pouvoir des alternatives. Cependant, comme nous l'avons vu dans le chapitre I, il n'est pas certain que le *pouvoir* d'agir autrement suppose toujours la *possibilité que l'agent agisse autrement*. Lorsqu'elle décrit des actions *bonnes* que l'agent ne peut pas éviter d'accomplir, Wolf imagine des situations où l'agent est déterminé par son propre caractère ou tempérament. Or ces exemples posent un problème d'interprétation : on se demande si le déterminisme en question est de nature causale et s'il enlève à l'agent son pouvoir d'agir autrement. Si rien ne permet d'affirmer qu'il s'agisse d'un déterminisme causal, on peut supposer que même si l'agent ne peut agir autrement, il en garde le pouvoir.

Reprenons son exemple de la femme altruiste qui se jette à l'eau pour sauver un enfant de la noyade. Wolf imagine que cette femme, à cause de son caractère foncièrement bon, ne peut s'empêcher de secourir le garçon. Elle serait donc *incapable* de ne pas se jeter à l'eau pour le sauver. Mais tout d'abord, comme Wolf le fait d'ailleurs par endroits⁵¹, on peut distinguer la *possibilité* et la *capacité*. Le fait que cette femme ne *puisse* pas rester tranquillement sur la plage prouve-t-il qu'elle en soit réellement *incapable* ? Ce n'est pas certain. Après tout, il n'est pas difficile de ne rien faire, ou de s'allonger sur le sable. Si aucun obstacle, aucun défaut psychologique ou physique, et aucune cause, ne l'empêche de rester passivement sur la plage, pourquoi n'en aurait-elle pas le pouvoir ? Conformément à l'hypothèse que j'ai adoptée dans le premier chapitre, je continue à admettre qu'il n'y a pas d'opposition claire entre la responsabilité morale et la *nécessité*, mais qu'il y en a une entre la responsabilité morale et la *nécessité causale*. En conséquence, il me semble que le principe des possibilités alternatives est vrai lorsqu'il est pris dans son sens causal, et douteux lorsqu'il est pris en un sens plus large. Dans tous les cas, même lorsque l'agent fait une action bonne, sa responsabilité est liée au fait qu'il lui était *causalement possible* d'agir autrement (ou de se retenir d'agir). Autrement dit, la responsabilité suppose toujours un *pouvoir* des alternatives. Mais il se pourrait tout de même qu'il y ait une asymétrie entre les actions bonnes et les

⁵¹ Cf. par exemple Suzan Wolf, *Freedom within Reason*, p. 100.

actions mauvaises. En effet, lorsqu'un agent accomplit une action bonne, le fait qu'il ne *puisse pas* agir autrement ne signifie pas nécessairement qu'il n'en ait pas le *pouvoir*.

Dans les cas où une action bonne est clairement déterminée par un facteur causal tel que l'hypnose, l'agent n'est manifestement pas responsable de son action. Wolf considère alors que l'action ne découle pas du moi profond de l'agent. Mais la thèse de Wolf est en difficulté lorsque l'action (bonne) découle du moi profond de l'agent, tout en résultant indirectement d'une manipulation, par exemple, si son moi profond a été modifié artificiellement. Ce genre de cas est analogue à son exemple « Jojo », sauf sur un point : dans l'exemple de Wolf, les actions de Jojo sont mauvaises. Qu'en est-il si l'agent fait le bien ? Supposons qu'on implante dans son cerveau un appareil produisant des inclinations altruistes. Après l'opération, il travaille pour les pauvres, rend visite aux prisonniers, participe au soutien scolaire dans un quartier défavorisé. Cet individu est-il responsable de ses actions sous prétexte qu'il agit selon son moi profond et possède la capacité d'agir selon le Vrai et le Bien ? Il semble que non. Le double critère proposé par Suzan Wolf est donc insuffisant.

La meilleure réponse à cette objection est peut-être que cet exemple diffère de « Jojo » par la nature des facteurs qui modifient le moi profond de l'agent. Dans l'exemple de Wolf, l'agent ne subit aucune transformation physique, mais est influencé selon un processus mimétique parfaitement naturel dans l'éducation des enfants. Comme les autres enfants, le jeune garçon se fie à ses parents et imite spontanément le comportement de son père. Dans un exemple strictement analogue à « Jojo » mais où l'agent fait le bien, ses bonnes actions viennent d'un moi profond et sain formé selon les mêmes processus que dans « Jojo ». Or dans ce cas, l'agent semble bien responsable de ce qu'il fait. Cette comparaison suggère qu'il y a bien une asymétrie entre les bonnes et les mauvaises actions du point de vue des conditions de la responsabilité morale. Néanmoins, l'exemple précédent, où le moi de l'agent est modifié artificiellement, montre que les deux conditions énoncées par Wolf ne suffisent pas. Il faut y ajouter une troisième condition relative à la continuité du « moi profond » ou aux capacités qu'il doit posséder pour être « sain ». Suffit-il de dire que le moi profond de l'agent ne doit pas avoir été modifié artificiellement ? Sans doute faut-il montrer aussi pourquoi les modifications artificielles de son moi profond excluent par la suite sa responsabilité morale. Ici, je suis tenté de dire, premièrement que seules certaines modifications artificielles

l'excluent, et deuxièmement, qu'elles l'excluent parce qu'elles enlèvent à l'agent la *possibilité causale* d'agir autrement. L'influence bienfaisante que constitue une bonne éducation morale respectueuse de la liberté de son destinataire ne diminue pas, normalement, son *pouvoir* de faire le mal, même si elle diminue la *probabilité* qu'il le fasse. Mais si la probabilité qu'il agisse mal diminue à cause d'un procédé qui lui enlève le pouvoir de le faire en certaines occasions, l'agent n'est pas responsable de ce qu'il fait dans ces situations.

Quoi qu'il en soit, ces réflexions montrent bien l'insuffisance d'une théorie du « moi profond », c'est-à-dire d'une théorie compatibiliste selon laquelle l'agent est responsable de son action lorsque celle-ci exprime sa personnalité profonde, ses valeurs, ou ses préférences. En effet, cette idée est incapable de rendre compte des cas où le moi profond de l'agent, indépendamment de sa volonté, a été modifié dans son contenu (valeurs, désirs, etc.) ou a perdu des capacités. Pour rendre compte de ces cas, le compatibilisme a besoin d'énoncer, comme Wolf, Fischer et Ravizza, des conditions supplémentaires relatives aux capacités de l'agent. Dans le prochain chapitre, l'étude de la responsabilité dans les cas d'omissions nous montrera mieux encore pourquoi le compatibiliste doit supposer, comme le libertarien, que la responsabilité de l'agent dépend étroitement de ses capacités.

4.9 Résumé

La responsabilité morale est liée aux pouvoirs de l'agent, et les compatibilistes eux-mêmes veulent donner une place à l'idée d'un pouvoir des alternatives, bien qu'ils comprennent ce pouvoir en un sens plus faible que le sens libertarien. L'analyse conditionnelle de ce pouvoir permet de lui donner une place dans une vision déterministe des actions humaines. Mais cette analyse est insuffisante si elle n'est pas appliquée en même temps aux désirs de l'agent. Or un tel contrôle suppose lui aussi un pouvoir des alternatives. Le compatibilisme est ainsi conduit à envisager une hiérarchie de désirs : pour que l'agent exerce un contrôle conscient sur ses désirs, il faut qu'il les veuille ou les approuve. Il a donc des désirs au sujet de ses désirs.

L'analyse conditionnelle des pouvoirs et l'idée d'une hiérarchie de désirs permet aussi de comprendre *l'autonomie* de manière compatibiliste : un agent est autonome s'il exerce un certain contrôle sur ses désirs à partir de son « moi profond ». Cependant, comme nous l'avons vu dans les sections 4.5 et 4.6 en appliquant l'argument du transfert, il ne peut s'agir d'une autonomie ultime. D'un point de vue compatibiliste, il est impossible que nous ayons la responsabilité ultime de nos actions. Pour cette raison, Wolf a choisi de centrer sa théorie non pas sur l'autonomie de l'agent, mais sur son aptitude à la rationalité. Selon elle, il ne suffit pas de dire que l'agent est responsable de ses actions lorsqu'elles découlent de son moi profond. Il faut aussi que son moi soit « sain », c'est-à-dire suffisamment apte à faire des choix rationnels.

Dans les chapitres V et VI, j'aurai l'occasion d'examiner en détail cette idée sous une forme plus élaborée que chez Wolf. Je m'intéresserai particulièrement à la théorie de Fischer et Ravizza, qui analyse les pouvoirs de l'agent en termes de mondes possibles. Dans le chapitre V, j'aborderai les cas d'omissions, car la responsabilité dans les omissions pose de sérieuses difficultés au compatibilisme en général. Après avoir critiqué deux réponses à l'objection des omissions, je me concentrerai sur une troisième réponse, la meilleure possible, à mon avis. Cette réponse se fonde sur l'analyse des pouvoirs en termes de mondes possibles. Dans le chapitre VI, je ferai une série d'objections contre cette réponse, et je montrerai que ces objections ne s'appliquent pas seulement à la responsabilité dans les omissions, mais aussi à la responsabilité dans les actions.

V Les omissions et le pouvoir des alternatives

5.1 Introduction

Les conditions de la responsabilité dans les omissions sont un sujet fort délicat pour les théories compatibilistes. Lorsqu'un individu est en procès pour non-assistance à une personne en danger, on ne lui reproche pas d'avoir accompli une action mauvaise ou d'avoir provoqué un événement malheureux, mais de ne pas avoir agi dans une situation où il devait le faire. Ainsi, la responsabilité à l'égard d'un événement ne suppose pas toujours d'accomplir une action qui provoque ou facilite cet événement. Ce qui est frappant est donc que l'agent responsable d'un événement par omission ne provoque pas un processus causal aboutissant à cet événement. Il est responsable pour n'avoir rien fait contre. Cela suggère qu'il pouvait faire quelque chose contre cet événement, et donc qu'il avait la possibilité d'agir autrement ou d'agir tout court. Ainsi, le cas des omissions milite en faveur d'un principe voisin de PAP :

(PPA - principe d'action possible) S est responsable de E par omission si E s'est produit et si S avait la possibilité d'agir efficacement contre E.

Face à cette conclusion, les compatibilistes ne sont pas nécessairement réduits à abandonner leur position en adoptant une version forte du principe des possibilités alternatives. Plusieurs solutions s'offrent à eux. Après avoir esquissé une théorie libertarienne de la responsabilité dans les cas d'omissions, je l'affinerais en abordant plusieurs difficultés dans les sections 5.3 à 5.5, puis je passerai à un examen systématique des solutions compatibilistes. Je pense que leur nombre est nécessairement très limité. La première solution compatibiliste, que je discuterai dans les sections 5.6 et 5.7, est de réduire les omissions à des actions. Cette solution se heurte à deux problèmes : tout d'abord, il y a des omissions qui semblent irréductibles aux actions parce qu'elles ne découlent pas d'une action ni même d'une décision. Deuxièmement, dans nos jugements intuitifs sur la responsabilité morale, il y a une

asymétrie entre les cas d'actions et les cas d'omissions, or la réduction des omissions aux actions ne tient pas compte de cette asymétrie.

La deuxième solution compatibiliste (que j'examinerai dans la section 5.8) est d'admettre le principe PPA, mais en soutenant que le *pouvoir des alternatives* requis par la responsabilité dans les omissions peut s'analyser de manière *conditionnelle*, sans supposer que l'agent a une *possibilité alternative* au sens le plus fort. Enfin, une troisième solution défendue par Fischer et Ravizza (et que j'examinerai seulement dans le chapitre VI) est d'analyser ce pouvoir en termes de *mondes possibles*, en soutenant qu'un agent est responsable d'un événement par omission s'il y a un monde possible proche du monde actuel où il fait quelque chose d'efficace contre cet événement. Ces deux dernières solutions se fondent sur une compréhension du pouvoir des alternatives qui ne s'oppose pas au déterminisme. Mais je soutiendrai qu'elles posent également problème, parce qu'elles ne sont pas toujours compatibles avec nos jugements intuitifs sur la responsabilité des agents, et qu'elles font nécessairement appel aux « conditions normales de délibération » sans pouvoir les définir.

5.2 Une théorie libertarienne de la responsabilité dans les omissions

Les omissions peuvent être l'objet de conflits d'intuitions, mais la plupart des philosophes admettent que dans certains cas d'omissions, la responsabilité est liée à un pouvoir des alternatives. On peut le voir dans ces deux exemples :

[Noyade] Jean se promène sur la plage. Il aperçoit soudain un enfant qui se débat dans l'eau pour échapper à la noyade. Jean délibère en lui-même un bref instant, puis continue négligemment sa promenade, et l'enfant se noie.¹

[Requins] Jean se promène sur la plage. Il aperçoit soudain un enfant qui se débat dans l'eau pour échapper à la noyade. Jean délibère en lui-même un bref instant, puis continue négligemment sa promenade, et l'enfant se noie. Or à l'insu de Jean, il y a des requins aux alentours. S'il s'était jeté à l'eau pour sauver l'enfant, il n'aurait pu le sauver car les requins l'auraient attaqué.

¹ Ces deux exemples viennent de Harry Frankfurt, « An Alleged Asymmetry between Actions and Omissions », *Ethics* 104 (1994), p. 620.

On s'accorde en général pour dire que dans « Noyade », Jean est responsable de la noyade, alors qu'il ne l'est pas dans « Requins ». On s'accorde aussi sur le fait que s'il n'est pas responsable de la noyade dans le deuxième cas, c'est parce que s'il avait essayé de sauver l'enfant, il n'aurait pu le faire. Nous pouvons donc poser le principe suivant :

(RO) (responsabilité dans une omission) : Un agent S est responsable par omission de l'événement E seulement s'il y a au moins une action A telle que :

- (a) S ne fait pas A, et E se produit.
- (b) S peut faire A.
- (c) Si S faisait A, E ne se produirait pas
- (d) S sait qu'il peut faire A, et que s'il faisait A, E ne se produirait pas.
- (e) S sait que E doit être évité.

Les conditions (d) et (e) sont les conditions épistémiques de la responsabilité. A la fin du chapitre III, j'ai énoncé des conditions du même genre pour les actions². On n'est pas responsable d'une action dont ignore complètement la valeur. Par exemple, si un enfant cueille des fleurs pour sa mère dans un parterre du château de Versailles, il n'est pas moralement responsable de ce qu'il fait ou provoque, car il ne sait pas qu'une telle action est interdite. Cela vaut aussi pour les omissions. Imaginons un individu qui ne sait pas qu'il doit s'incliner devant son roi lorsque celui-ci passe devant ses sujets. S'il omet de s'incliner, il n'est pas responsable de cette omission. On peut se demander s'il ne vaudrait pas mieux formuler la condition (e) de la même manière que pour les actions, ce qui donnerait : « S sait si E est plutôt bon ou mauvais ». On tiendrait compte ainsi des omissions *méritoires*. Mais il ne semble pas qu'on puisse être responsable d'un fait bon ou heureux par omission.

Par exemple, imaginons le cas suivant :

[dénonciation] Pierre déteste Madeleine. Or Madeleine a eu des relations amoureuses avec deux officiers allemands pendant la guerre. A la libération, Pierre pourrait dénoncer cette femme et la soumettre ainsi à la vindicte populaire. Il y pense avec une

² Cf. *supra*, p.169 et s.

certaine envie, puis trouve cette action lâche et injuste. Il décide donc de ne pas la dénoncer, et personne ne dénonce Madeleine.

Pierre est-il responsable du fait que Madeleine n'ait pas été humiliée en public à la fin de la guerre ? Bien que nous puissions le louer pour son choix de ne pas dénoncer Madeleine, surtout s'il a été gravement offensé par elle, il ne semble pas responsable du fait qu'elle ait échappé à la vindicte populaire. En revanche, lorsqu'un individu a laissé un événement heureux se produire grâce à une omission héroïque ou extrêmement méritoire, on se demande s'il n'en est pas responsable, par exemple s'il a résisté à des tortures atroces pour éviter de dénoncer quelqu'un. Dans ce cas, le héros en question a-t-il sauvé la vie de la personne qu'il aurait pu dénoncer ? Est-il responsable du fait qu'elle soit encore en vie ? Il est difficile de répondre à cette question. Dans le doute, je ferai comme si la responsabilité dans les omissions s'applique seulement aux événements malheureux, sachant que la réponse à cette question n'a pas d'incidence sur les questions abordées dans ce chapitre. En ce qui concerne les événements malheureux, nous pouvons remarquer que la responsabilité dans les omissions existe seulement par rapport à des événements malheureux qu'on a le *pouvoir* et le *devoir* d'éviter. Si nous pouvions être responsable par omission des autres sortes d'événements, nous porterions la responsabilité d'une infinité d'événements, car il y a une infinité de choses que nous ne faisons pas mais que nous pourrions faire si nous le voulions. Beaucoup d'actions ne sont pas obligatoires mais *surérogatoires*. Par exemple, je pourrais rendre visite aux malades, mais aussi aux prisonniers, aux personnes âgées, aux SDF, et à des personnes qui souffrent de la solitude. Je pourrais aussi faire des dons à une multitude d'associations caritatives. Si je ne fais rien de tout cela, suis-je responsable de tous les événements malheureux contre lesquels j'aurais pu agir ? Non, sinon un même individu pourrait être responsable de toutes les formes de pauvreté existant dans son pays. La responsabilité par omission ne concerne donc que les événements qu'on a le pouvoir et le devoir d'éviter³. D'où la formulation utilisée ici pour la condition (e).

³ Notons cependant qu'il y a sans doute un devoir collectif de lutter contre toutes ces formes de pauvreté à la fois. Une nation ou un peuple peut donc être responsable par omission d'une grande partie de la pauvreté qui existe en son sein.

La condition épistémique (d) est aussi importante que la (e). Par exemple, supposons que Jean se trouve dans la même situation que dans « Requins », sauf qu'il voit les requins, et qu'à son insu, un hélicoptère et une équipe de sauveteurs se trouve à vingt mètres de lui, derrière la dune. Il peut sauver l'enfant en appelant les sauveteurs et en leur indiquant où se trouve l'enfant. Mais il ne sait pas qu'il peut faire cela. Il veut sauver l'enfant et cherche désespérément comment le sauver, mais ne voit aucun moyen d'y parvenir. Il le laisse donc se noyer.

Jean avait un moyen de sauver l'enfant et il ne l'a pas sauvé. Pourtant, il n'est pas responsable de sa mort. Cela s'explique par le fait qu'il ne connaissait pas ce moyen. D'où l'importance de (d). Dans les actions, la condition (d) est remplacée par celle-ci :

S sait qu'il fait A et qu'en faisant A, il provoque E ou contribue à E.

Les conditions épistémiques de la responsabilité ne sont pas ce qui oppose les compatibilistes et les incompatibilistes. Le point le plus délicat est la condition (a), que les libertariens comprennent au sens le plus fort : A doit être possible et accessible à S *compte tenu des lois de la nature et du passé actuel*. Indépendamment de ce débat, sur lequel je reviendrai plus loin, RO peut rencontrer plusieurs difficultés. La première difficulté est que dans les exemples du même genre que « Requins », l'agent semble avoir une certaine responsabilité morale. En effet, puisqu'il n'a rien fait pour sauver l'enfant, il est blâmable. Il ne pouvait pas le sauver, mais il pouvait tout de même *essayer* de le sauver. S'il avait vu les requins, il serait sans doute excusable, mais puisqu'il ne les a pas vus, il aurait dû essayer de sauver l'enfant. Faut-il en conclure que Jean est responsable de la noyade ? Non. Il est responsable de son choix (de ne pas nager vers l'enfant), et de sa passivité. Cela peut suffire pour rendre compte du caractère blâmable de son comportement.

On peut donc supposer que dans les cas d'omissions, la responsabilité de l'agent s'étend aussi loin que son pouvoir d'agir autrement, et c'est pour cette raison qu'il est blâmable : il est responsable de son mauvais choix et de son comportement passif, c'est-à-dire d'une omission consistant à *ne pas essayer de sauver* l'enfant. En revanche, il n'est pas responsable de la noyade, ni de l'omission consistant à *ne pas sauver* l'enfant.

Il ne semble pas nécessaire, d'ailleurs, d'énoncer des conditions différentes pour la responsabilité à l'égard d'un choix, d'une omission, et d'un événement. Dans RO, nous pouvons considérer que E désigne soit un effet de l'omission, soit le fait que S ne fasse pas A (donc l'omission elle-même), soit le fait qu'il ne choisisse pas de faire A (ce qu'on pourrait appeler une omission mentale). Par exemple, si dans « Requins », on remplace les requins par un intervenant contrefactuel qui contrôle les mouvements de Jean sans contrôler ses choix, Jean ne peut ni sauver l'enfant, ni essayer de le sauver en se jetant à l'eau. Il peut seulement décider de le sauver. Comme dans « Requins », Jean ne paraît pas responsable de la mort de l'enfant, mais il n'est pas plus responsable du fait qu'il ne se soit pas jeté à l'eau pour aller vers l'enfant. Il est seulement responsable d'un mauvais choix, celui de ne pas sauver l'enfant. Donc *la responsabilité de Jean s'étend aussi loin que son pouvoir des alternatives*. Appelons « E » le fait qu'il soit resté sur la plage. S'il n'y avait pas d'intervenant contrefactuel, Jean serait responsable de E car :

- (a) Il pourrait faire A (choisir de sauver l'enfant).
- (b) S'il faisait A, E ne se produirait pas (donc Jean ne resterait pas sur la plage).
- (c) Il sait qu'il peut faire A et que s'il faisait A, E ne se produirait pas.
- (d) Il sait si E est plutôt bon ou mauvais.

Les mêmes conditions restent valables si E désigne le choix de Jean de ne pas sauver l'enfant. Par exemple, on peut dire que Jean est responsable de E (ainsi défini) car :

- (a) Il aurait pu faire A (devenir courageux et altruiste à un degré normal)
- (b) S'il avait fait cela, E ne se produirait pas (Jean ne choisirait pas de ne pas sauver l'enfant).
- (c) Il savait qu'il pouvait devenir courageux et altruiste à un degré normal, et que s'il le faisait, E ne se produirait pas.
- (d) Il savait que E devait être évité.

Dans cet exemple, nous voyons que les conditions (c) et (d) sont trop fortes. Il suffit que l'agent *puisse savoir* qu'il peut faire A, et *puisse connaître* la valeur de A et de ses conséquences. Par exemple, pour que Jean soit responsable de son choix, il suffit qu'il ait pu savoir qu'un tel choix est mauvais et qu'en devenant courageux et altruiste à un degré normal,

on évite de faire ce genre de choix. L'ignorance du bien et du mal, et même l'ignorance tout court, est parfois coupable : lorsqu'elle a des conséquences mauvaises et que l'agent a le pouvoir de l'éviter. C'est pourquoi RO doit être modifié de la façon suivante :

(RO*) Un agent S est responsable par omission de l'événement E seulement s'il y a (au moins) une action A telle que :

- (a) S ne fait pas A, et E se produit.
- (b) S peut faire A.
- (c) Si S faisait A, E ne se produirait pas.
- (d) S *peut savoir* qu'il peut faire A et que s'il faisait A, E ne se produirait pas.
- (e) S *peut savoir* que E doit être évité.

Enfin, on peut ajouter :

- (f) S a le devoir de connaître la valeur de E, et il sait qu'il a ce devoir.

Cette dernière condition est importante, car si S n'a pas le devoir de se renseigner sur la valeur de E, on ne peut pas lui reprocher ensuite de ne pas avoir voulu empêcher E. Dans la compréhension libertarienne de ces conditions, le mot « peut » dans les conditions (c), et (d), se comprend comme dans (a), c'est-à-dire au sens le plus fort : il se réfère à un événement possible compatible avec les lois de la nature et le passé actuel, et accessible à S.

5.3 Première difficulté. La distinction entre les omissions et les actions

Une première question au sujet des omissions est de savoir si elles se distinguent nettement des actions. Forment-elles une sous-catégorie parmi les actions ? Y a-t-il une différence nette entre les deux ou seulement une distinction de degrés ? Et si la différence est nette, peut-on penser que l'action et l'omission sont deux aspects complémentaires d'un même phénomène ? Cette troisième hypothèse est attirante, car sans nier l'asymétrie entre l'action et l'omission, elle implique que les conditions de la responsabilité morale sont rigoureusement les mêmes dans toutes les situations. Ces conditions seraient seulement susceptibles d'être

décrites de deux manières différentes selon que l'on considère le comportement de l'agent en tant qu'action ou en tant qu'omission. Prenons un exemple :

[Princesse 1] La princesse est à l'opéra. La première partie du concert se termine. Pendant l'entracte, la princesse reste d'abord dans sa loge, et selon la coutume, elle se lève et se tient debout pendant une minute, durée suffisante pour qu'elle soit regardée par ses sujets et photographiée pour les journaux. Or la princesse a envie de rester debout plus longtemps, faisant fi de l'étiquette. Elle se tient debout pendant quatre minutes.⁴

La princesse semble responsable de son comportement et de ses conséquences, par exemple du fait qu'elle soit photographiée pendant plusieurs minutes. Si nous décrivons son comportement comme une action (« elle se tient debout pendant quatre minutes »), ce jugement est confirmé, car il correspond aux conditions de la responsabilité dans une action. En effet, la princesse se tient debout pendant quatre minutes après l'avoir décidé dans des conditions normales, sans y être forcée ou déterminée causalement, en sachant ce qu'elle faisait et si son action était correcte ou non. Elle est donc responsable de son choix, de son action et des conséquences de son action. Et si nous décrivons ce comportement comme une omission (« La princesse omet de s'asseoir au bout d'une minute »), les conditions de la responsabilité dans les omissions aboutissent au même jugement : elle a fait cela alors qu'elle pouvait s'asseoir au bout d'une minute. De plus, elle savait qu'elle pouvait s'asseoir au bout d'une minute, qu'elle devait le faire, et qu'en le faisant, elle aurait évité d'être photographiée trop longtemps. C'est pourquoi elle est responsable du fait qu'elle a été photographiée pendant plusieurs minutes.

On pourrait donc croire que toute omission consiste en une action et que toute action A est en même temps l'omission consistant à ne pas se retenir de faire A. Cependant, dans les exemples d'omissions où l'agent ne peut pas agir autrement (à cause d'un obstacle ou d'un intervenant contrefactuel), il apparaît qu'on ne peut décrire le comportement de l'agent à la

⁴ J'emprunte cet exemple et le suivant à Alison McIntyre. Cf. A. McIntyre, "Compatibilists Could Have Done Otherwise: Responsibility and Negative Agency". *The Philosophical Review*, Vol.103/3 (1994), p. 453-488. Ces deux exemples se trouvent p. 474-475.

fois comme une action et une omission, car ces deux descriptions aboutissent à des jugements opposés. C'est le cas ici :

[Princesse 2] Comme dans l'exemple précédent, la princesse se tient debout pendant quatre minutes, bravant la coutume selon laquelle elle doit s'asseoir au bout d'une minute. Mais elle ne peut s'asseoir au bout d'une minute car un ingénieur surdoué payé par les photographes contrôle ses mouvements au moyen d'un champ magnétique, si bien qu'elle peut à peine changer de pose pendant quatre minutes.

Si nous décrivons le comportement de la princesse comme une action, elle est responsable, comme dans l'exemple précédent, car dans les cas d'actions, la présence d'un intervenant contrefactuel ne retire pas à l'agent sa responsabilité morale. La princesse est donc responsable de son choix et de tout ce qui en découle : l'action et ses conséquences. Mais si nous voyons le comportement de la princesse comme une omission, la conclusion n'est plus la même, car elle ne peut s'asseoir au bout d'une minute, ni éviter, par conséquent, d'être photographiée pendant plusieurs minutes. Elle n'est donc pas responsable de son omission, ni du fait qu'on la photographie pendant plusieurs minutes.

Puisque les deux descriptions sont incompatibles entre elles, nous devons supposer que le comportement de la princesse n'est pas vraiment une action et une omission à la fois, ou bien corriger RO*. Je pense que la meilleure solution est la première. Le comportement de la princesse est une action et non une omission, car en se tenant debout pendant plusieurs minutes, elle n'omet pas seulement de se dérober aux regards de ses sujets, mais les incite à la regarder pendant tout ce temps.

Nous pouvons donc énoncer ce critère :

(DAO) (différence entre l'action et l'omission) :

1. Un comportement de S qui provoque ou facilite l'événement E est une *action* à l'égard de E.
2. Un comportement de S qui ne provoque ni ne facilite l'événement E, mais qui ne s'y oppose pas non plus, est une *omission* à l'égard de E.

Par exemple, un trafiquant d'armes qui fournit de bonnes armes aux anglais pendant la guerre de cent ans participe indirectement aux victoires anglaises et aux défaites françaises. Ce cas est donc un cas d'action. S'il ne fournit pas d'armes aux anglais ni aux français, il *omet* seulement de s'opposer aux victoires anglaises. Et dans ce cas, il omet aussi de s'opposer aux victoires françaises⁵. Si l'exemple de la princesse paraît ambigu, c'est parce que pour savoir si un comportement est une action ou une omission de la part de l'agent, il faut tenir compte de la réactivité de son environnement. Or dans le cas de la princesse, cette réactivité est très grande : un mot ou un geste de la princesse suffit à provoquer une réaction importante de la foule, de la presse et des médias. Si les photographes n'interprétaient pas le comportement de la princesse comme une invitation à la photographier plus longtemps, ce comportement ne serait plus qu'une omission à l'égard du fait qu'on la prenne en photo pendant plusieurs minutes.

Inversement, une activité intense doit être vue comme une omission par rapport à E si en se livrant à cette activité, l'agent ne fait rien qui provoque ou facilite E. Par exemple, si Jean fait des pompes sur la plage au moment où l'enfant se noie, il fait une action, mais il omet de sauver l'enfant.

Ainsi, on voit que du point de vue de la responsabilité morale, le caractère « actif » du comportement de l'agent ne dépend pas de sa mobilité ou de l'énergie qu'il déploie, mais de la relation qui existe entre ce comportement et l'événement en question. Un même comportement peut être à la fois une action et une omission, mais seulement par rapport à deux événements ou états de choses différents. Par exemple, en faisant la fête juste avant ses examens, un étudiant peut à la fois contribuer activement à la bonne ambiance de son groupe d'amis, et omettre de réviser pour ses examens⁶. En somme, puisque l'omission n'est pas nécessairement une absence d'activité, la notion d'omission est *relative*.

⁵ S'il donne des armes à la fois aux français et aux anglais, il est responsable de deux actions à la fois, et il peut ainsi être partiellement responsable des victoires françaises et des victoires anglaises. Mais bien sûr, la victoire finale n'appartient qu'à une des deux armées. Il est donc partiellement responsable de cette victoire finale, et par là, de la défaite finale de l'autre armée. Sa part de responsabilité dépend bien sûr de la quantité et de la qualité des armes qu'il aura vendues au vainqueur. Sur ce genre de cas, cf. aussi note suivante.

⁶ À ce propos, notons qu'une même action ou une même omission peut être favorable à E sous un certain aspect, et favorable à non-E sous un autre. Par exemple, dans un contexte de guerre entre les français et les anglais, un

5.4 Deuxième difficulté : l'omission peut-elle causer un effet ?

Une deuxième difficulté au sujet des omissions concerne cette relation entre le fait de ne pas faire A et l'événement E dont on est considéré comme responsable *parce qu'on n'a pas fait* A. Quelle est la nature de cette relation ? Est-ce une relation de causalité ?

A première vue, l'omission ne cause pas E, car il est impossible d'identifier un processus physique allant de l'omission à E. Par exemple, dans « Noyade », où Jean est responsable de la mort de l'enfant en restant sur la plage alors qu'il a la possibilité de le sauver, il n'y a aucun processus allant de « Jean reste sur la plage à t1 » à « l'enfant se noie à t2 ».

On peut supposer que la relation entre E et O (l'omission) n'est pas une relation causale mais une relation de dépendance contrefactuelle. Avant d'examiner cette hypothèse, rappelons ce qu'est la dépendance contrefactuelle.

(déf.) Y dépend contrefactuellement de X si et seulement si, si X ne s'était pas produit, Y ne se serait pas produit non plus.⁷

L'hypothèse que je veux maintenant examiner est que le transfert de responsabilité de O à E se fait selon le principe :

(TR contrefactuel) La responsabilité d'un agent à l'égard de Y se transmet de X à Y si et seulement si 1°) Y dépend contrefactuellement de X, et 2°) les conditions épistémiques sont satisfaites.

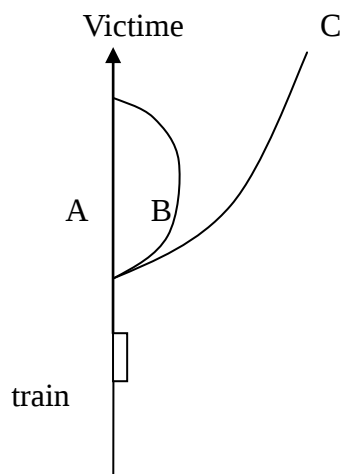
breton peut décider de ne soutenir personne. Il omet ainsi à la fois de soutenir les anglais et de soutenir les français. Il peut aussi faire une vente d'armes qui profite à la fois aux deux armées, ce qui est une action favorable aux victoires anglaises et aux victoires françaises. Mais bien sûr, puisque la victoire à une même bataille n'appartient qu'à une seule armée, on ne peut pas causer par une action (ou par une omission) à la fois la victoire des anglais, et celle des français. Donc si une action ou une omission peut sembler favorable à E et à non-E à la fois, elle ne peut pas *causer* E et non-E à la fois, car ces deux événements sont logiquement incompatibles.

⁷ Je reprends ici les définitions utilisées par Carolina Sartorio dans « A New Asymmetry Between Actions and Omissions », *Noûs* 39/3 (2005), p. 460-482. Sartorio pose ces définitions p. 466.

Par exemple, dans « Noyade », si Jean n'avait pas décidé de ne pas sauver l'enfant, il l'aurait sauvé, donc l'enfant ne se serait pas noyé. La noyade dépend donc contrefactuellement de l'omission consistant à ne pas essayer de sauver l'enfant, et Jean est responsable de la noyade en étant responsable de cette omission. Au contraire, dans « Requins », Jean n'est pas responsable de la noyade, bien qu'il soit responsable de son omission, et cela s'explique par le fait que la noyade ne dépend pas contrefactuellement de l'omission. Enfin, si on remplace les requins par un savant fou qui empêche Jean de sauter à l'eau sans l'empêcher de choisir de sauver l'enfant, la mort de celui-ci ne dépend pas contrefactuellement de l'omission de Jean, et celui-ci, en outre, n'est pas responsable de son omission puisqu'il ne pouvait même pas *essayer* de sauver l'enfant (il pouvait seulement *décider* de le sauver).

Carolina Sartorio a imaginé un exemple pour réfuter TR contrefactuel, mais cet exemple est particulièrement difficile à interpréter :

[Aiguillage] Vous vous trouvez le long d'un chemin de fer, à côté d'une bifurcation à trois branches. La branche A se dirige vers un homme ligoté sur les rails, et la branche B rejoint la voie A juste avant l'endroit où se trouve l'homme ligoté. Seule la voie C ne passe pas par cet endroit. Vous savez tout cela. Un train arrive et l'aiguillage est orienté vers la voie A. Vous hésitez entre deux actions : orienter l'aiguillage vers la voie C, et l'orienter vers la voie B. Finalement, vous avez envie de l'orienter vers la voie B. Le train écrase donc l'homme ligoté sur la voie. Or à votre insu, un neurochirurgien contrôlait vos mouvements de telle sorte que même si vous aviez choisi d'orienter l'aiguillage vers la voie C, il vous aurait fait placer l'aiguillage vers la voie B.⁸



⁸ *Ibid.*, p. 467.

Sartorio interprète l'exemple de la manière suivante :

1. Vous êtes responsable d'avoir orienté l'aiguillage vers la voie B, car c'est librement que vous avez choisi de l'orienter dans cette direction.
2. Si vous n'aviez pas orienté l'aiguillage vers la voie B, vous l'auriez dirigé vers la voie C et l'homme ne serait pas mort. La mort de l'homme ligoté sur les rails dépend donc contrefactuellement du fait que vous ayez dirigé l'aiguillage vers la voie B.

Si TR contrefactuel est vrai, il découle de (1) et de (2) que vous êtes responsable de la mort de l'homme ligoté sur les rails. Or vous n'en êtes pas responsable, car si vous aviez voulu le sauver, vous n'auriez pu le faire : le neurochirurgien vous en aurait empêché.

Sartorio conclut que TR contrefactuel est faux. Mais je pense que son interprétation de l'exemple n'est pas correcte. En effet, elle le traite à la fois comme un cas d'action et comme un cas d'omission. Comme nous l'avons vu plus haut⁹, pour savoir, dans un exemple à la Frankfurt, si un agent est responsable d'un événement E, il faut savoir si son comportement à l'égard de E est une action ou une omission. Autrement dit, il faut savoir si son comportement n'empêche pas E, ou s'il provoque (ou facilite) E. Dans le premier cas, il s'agit d'une omission, et dans le second, d'une action. Et ce critère est admis par Sartorio elle-même.

En dirigeant l'aiguillage vers la voie B, vous semblez d'abord faire une action, mais cette impression est erronée, car faire cela équivaut à ne rien faire. En agissant ainsi, vous ne diminuez pas la menace qui pèse sur l'homme ligoté, mais vous ne l'augmentez pas non plus. Si votre comportement avait créé ou augmenté la menace, il faudrait le considérer comme une action, mais puisque vous n'avez pas augmenté cette menace, votre action est en fait une omission à l'égard du résultat (la mort de l'homme). Si le neurochirurgien ne contrôlait pas vos mouvements, on pourrait vous reprocher de ne pas avoir dirigé l'aiguillage vers la voie C. Mais à cause du neurochirurgien, vous n'êtes pas responsable de cette omission, car vous

⁹ Cf. *supra*, p. 223 et s.

n'auriez pas pu placer l'aiguillage vers la voie C. Ainsi compris, l'exemple ne contredit plus TR contrefactuel car :

- la mort de l'homme dépend contrefactuellement de votre omission (qui consiste concrètement dans le fait de placer l'aiguillage sur B).
- vous n'êtes pas responsable de la mort de l'homme ligoté sur les rails.
- Mais vous n'êtes pas non plus responsable de votre omission.

Si le fait de placer l'aiguillage sur B était une action, vous en seriez responsable, car dans les exemples d'actions (comme nous l'avons vu au chapitre II), la présence d'un intervenant contrefactuel qui empêche l'agent d'agir autrement n'empêche pas celui-ci d'être responsable de son action¹⁰. Mais il n'en est pas de même dans les cas d'omissions, comme l'admet d'ailleurs Sartorio. Selon Sartorio, en effet, un comportement volontaire qui ne diminue pas une menace déjà présente mais ne l'augmente pas non plus est une omission. Sartorio devrait donc considérer comme une omission le fait de placer l'aiguillage vers B. Ainsi, elle devrait raisonner à propos de cette omission de la même manière qu'à propos de l'omission consistant à ne pas sauver l'homme ligoté sur les rails : dans un cas comme dans l'autre, vous n'êtes pas responsable. C'est pourquoi le contre-exemple imaginé par elle ne fonctionne pas, et TR contrefactuel demeure intact. Pour la suite, je supposerai donc que ce principe est correct, mais en le modifiant légèrement pour rendre compte de certains exemples. Supposons que si je joue du piano, cela implique que je n'ai pas la maladie xyz, Dans ce cas, le fait que j'aie la maladie xyz dépend contrefactuellement du fait que je joue du piano. Donc si TR contrefactuel est vrai, alors si je suis responsable du fait que je joue du piano, je suis aussi responsable du fait que je n'ai pas la maladie xyz, ce qui est évidemment absurde. Dans cet exemple, Y est antérieur à X. Nous pourrions donc corriger TR contrefactuel en précisant que Y doit être postérieur à X, mais cela suffirait-il ?

¹⁰ L'intervenant contrefactuel n'élimine pas tout *flicker of freedom* (c'est pourquoi on peut défendre le principe des possibilités alternatives pour les actions), mais il rend inévitable une certaine action *générale*, par exemple « Gunnar tue Ridley à t ».

Carolina Sartorio pense que dans les cas d'omissions, la responsabilité ne se transmet pas à travers une relation contrefactuelle, mais à travers une relation causale¹¹. Elle propose donc d'utiliser le même principe de transfert que pour les actions :

(TR causal) La responsabilité d'un agent à l'égard de X se transmet à Y si et seulement si 1°) X cause Y, et 2°) les conditions épistémiques sont satisfaites.

Mais dans les cas d'omissions, X ne cause pas Y de la même manière que dans les actions, c'est-à-dire par l'intermédiaire d'un processus plus ou moins identifiable comme un mouvement, un courant électrique, ou (dans « Princesse ») un signe visuel donné par la princesse à ses sujets. Or la relation causale particulière qui relie l'omission à ses conséquences a un aspect contrefactuel. Je pense qu'on pourrait définir cette relation de la manière suivante :

(déf.) Y dépend *contrefactuellement et causalement* de X $\leftrightarrow$ si non-X, alors non-X causerait non-Y.

On pose alors le principe suivant :

(TR contrefactuel et causal) La responsabilité d'un agent à l'égard de X se transmet à Y si et seulement si 1°) Y dépend *contrefactuellement et causalement* de X, et 2°) les conditions épistémiques sont satisfaites.

Cela indique en quel sens une omission peut *causer* un effet et en quel sens TR causal s'applique aux cas d'omissions. Les omissions *causent* des effets au sens où le manque de vêtements chaud peut causer la mort par hypothermie ou au sens où le manque de pluie peut causer la sécheresse et une mauvaise récolte : le manque de pluie n'est pas une cause efficiente mais une cause *déficiente*. Au sens large, la cause d'un état de chose est ce que nous désignons pour expliquer cet état de chose (une substance, un événement, ou un état de chose peuvent être causes). En ce sens, une cause n'est pas nécessairement une cause efficiente. Un

¹¹ *Ibid.*, p. 468.

événement X est cause d'un événement Y si X est contraire à ce qui était attendu, et si non-X (l'événement attendu) aurait normalement causé non-Y.

Cette forme de causalité existe entre les choses inertes et parmi les êtres vivants non-rationnels. Il n'y a donc pas d'inconvénient à dire qu'on cause certains événements par omission. Seulement, cette causalité est d'un type particulier en raison de son aspect contrefactuel. Ainsi, TR causal s'applique aux omissions, mais au sens indiqué par TR contrefactuel et causal.

5.5 Troisième difficulté. Le débat sur la transitivité de la causalité

Si on pense que la responsabilité dans les omissions se transfère de manière causale, on rencontre une difficulté devant certains exemples ayant la structure suivante : un agent ne fait pas l'action A, et un certain résultat se produit. Si l'agent faisait A, le même résultat se produirait, mais en résultant d'une autre menace que celle ayant cours dans le monde actuel. Voici un cas de ce genre :

[Deux tueurs 1] Je vois Gunnar pointer son arme vers Ridley et je le laisse tirer. J'aurais pu l'en empêcher mais je ne l'ai pas fait. Ridley s'écroule et meurt quelques minutes plus tard. Or, à mon insu, un autre tueur visait Ridley. Si j'avais empêché Gunnar de tirer, ce tueur l'aurait abattu et le résultat aurait été le même.

Dans cet exemple, il semble que je ne sois pas responsable de la mort de Ridley, car même si j'avais essayé de le sauver, un tueur l'aurait abattu. Par contre, je suis responsable de mon omission, car j'aurais pu *essayer* de sauver Ridley. Or si dans la situation actuelle, mon omission cause d'une certaine manière la mort de Ridley, TR causal est faux, car bien que je sois responsable de mon omission et que celle-ci cause la mort de Ridley, je ne suis pas responsable de la mort de Ridley.

Si on pense que je ne suis pas responsable de sa mort, cet exemple ne pose pas de problème. Mais si on pense que j'en suis responsable, et que TR causal est vrai, il faut

remettre en question la transitivité de la causalité, c'est-à-dire le principe selon lequel si X cause Y et si Y cause Z, alors X cause Z.

En effet, en contestant ce principe, on peut soutenir ceci :

- Mon omission cause le fait que Ridley ne soit pas protégé contre Gunnar.
- Le fait que Ridley ne soit pas protégé contre Gunnar cause sa mort.
- Mais mon omission ne cause pas la mort de Ridley.

Notons que si la causalité n'était pas transitive, le transfert causal d'impuissance et le transfert causal de non-responsabilité que j'ai défendus dans le chapitre III deviendraient inutilisables. Mais il ne paraît pas nécessaire de remettre en question la transitivité de la causalité. L'exemple proposé est un cas d'omission. Donc si dans cet exemple, je ne suis pas responsable de la mort de Ridley, cela peut venir du fait que la causalité particulière à l'œuvre dans les cas d'omissions n'est pas transitive. Cela ne remet donc pas en question la transitivité de la causalité *efficiente*. Jusqu'à preuve du contraire, la causalité physique ou naturelle dont il s'agit dans l'argument du transfert doit être considérée comme transitive. Or je pense que si la causalité n'est pas transitive dans les cas d'omissions, c'est précisément à cause de son caractère *contrefactuel*.

Ici, comme dans la discussion sur « Betty » et le principe TNR¹², il faut distinguer l'événement-type et l'événement-token. Dans l'exemple « Deux tueurs 1 », je ne suis pas responsable de la mort de Ridley *en général*, car si j'avais essayé de le sauver, il serait mort, et si le deuxième tueur l'avait manqué, il serait mort aussi d'une manière ou d'une autre, car tout homme doit mourir un jour. Quand on attribue la responsabilité de la mort d'une personne à un individu, on n'envisage pas cet événement comme un *universel*, sinon personne ne pourrait être responsable de la mort de quelqu'un. Mais faut-il en conclure qu'il s'agit d'un *événement-token* ? Non plus, car dans cette hypothèse, je serais pleinement responsable de la mort de Ridley, même si je ne pouvais pas empêcher son assassinat.

¹² Cf. *supra*, p.154 et s.

Aussi, je pense que dans les cas de ce genre, la responsabilité de l'agent par rapport au résultat dépend du degré de différence qui sépare les deux événements comparés, c'est-à-dire l'événement actuel et l'événement alternatif qui se serait produit si l'agent était intervenu (et *malgré* cette intervention). Dans « Deux tueurs 1 », ma responsabilité paraît négligeable car les deux tentatives d'assassinat se situent à peu près au même moment, par exemple à 5 secondes d'intervalle. Mais s'il y a un an ou dix ans entre les deux tentatives (celle de Gunnar et celle de l'autre tueur), nous sommes tentés de dire que cette fois, je suis partiellement responsable de la mort de Ridley, car j'aurais pu l'empêcher. C'est que la mort de Ridley dans la situation actuelle ne semble pas être le même événement que la mort de Ridley dans la situation alternative. Or il n'y a qu'une différence de degrés entre cette différence et celle qui distingue les deux événements dans le cas précédent. Ainsi, plus les deux résultats comparés diffèrent par leur situation dans le temps et leur histoire causale, plus on est responsable du résultat actuel si on peut l'empêcher. Quand cette différence est minime, comme dans « Deux tueurs », la responsabilité de l'agent est négligeable. Mais quand la différence est beaucoup plus grande, comme dans cet exemple où l'histoire causale des deux événements n'est pas la même, l'agent porte une responsabilité non-négligeable :

[Deux tueurs 2] Gunnar veut tuer Ridley pour des raisons personnelles. Je le vois viser Ridley et je le laisse tirer. Ridley est touché et meurt en quelques minutes. J'aurais pu empêcher Gunnar de le tuer, mais si j'avais fait cela, Ridley aurait été tué dix ans plus tard par un agent du KGB qui l'aurait éliminé en sabotant sa voiture.

Ici, les deux résultats comparés diffèrent largement par leur situation temporelle et leur histoire causale. En conséquence, il paraît évident que je suis responsable de la mort de Ridley, même si Ridley devait de toute façon mourir un jour de mort violente.

Ces remarques indiquent une nouvelle asymétrie entre les actions et les omissions. Dans un cas d'*action* analogue à « Deux tueurs 1 », la responsabilité de l'agent ne dépend nullement du degré de différence qui sépare le résultat actuel du résultat alternatif. C'est ce que nous avons pu voir dans « Betty »¹³. Betty détruit le camp ennemi en provoquant une avalanche au

¹³ Cf. *supra*, p.154 et s.

moyen d'explosifs. Elle aurait pu ne pas appuyer sur le détonateur. Mais une brebis aurait alors heurté une pierre une seconde après, ce qui aurait provoqué une avalanche au même endroit, et le camp ennemi aurait été détruit. Dans cet exemple, Betty paraît pleinement responsable même si elle ne pouvait pas empêcher la destruction du camp ennemi. Au contraire, si Betty, au lieu d'appuyer elle-même sur le détonateur, avait *laissé* quelqu'un faire cela¹⁴, on ne la considérerait pas comme responsable de l'avalanche et de la destruction du camp ennemi. Et probablement, on ne dirait pas non plus qu'elle a *causé* ces événements. Je propose donc de considérer que dans « Deux tueurs 1 », non seulement je ne suis pas responsable de la mort de Ridley, mais en outre, mon omission n'a pas causé la mort de Ridley, car je ne pouvais rien faire qui eût permis d'aboutir à un résultat suffisamment différent.

Ces réflexions nous conduisent à poser le principe :

(ADD) (Asymétrie des degrés de différence) Pour tout événement particulier E :

- Si S cause E par une action, alors s'il pouvait faire que se produise à la place de E un événement E* ressemblant à E, S est responsable de E quel que soit le degré de ressemblance entre E et E*.
- Si S laisse E se produire, alors s'il pouvait faire que se produise à la place de E un événement E* ressemblant à E, la responsabilité de S à l'égard de E dépend du degré de différence entre E et E* (et si cette différence est négligeable, la responsabilité de S est négligeable).

Cela revient à dire que dans une omission, la responsabilité de l'agent à l'égard du résultat est proportionnée à ce qu'il pouvait faire pour obtenir un résultat différent. Cette différence s'explique par le fait que la causalité n'est pas la même dans les deux cas. Dans les actions, il s'agit de causalité *efficiente*, alors que dans les cas d'omissions, la relation causale a un aspect *contrefactuel* qui nous oblige à comparer l'événement en question avec celui qu'aurait provoqué l'agent s'il était intervenu. Cela explique enfin pourquoi la causalité n'est

¹⁴Cette expression ne signifie pas ici que Betty aurait ordonné à quelqu'un de le faire à sa place. Sinon, Betty serait encore responsable, dans ce cas, de l'avalanche et de la destruction du camp ennemi.

pas transitive dans les cas d'omissions. En effet, le résultat du comportement de l'agent peut être quasiment identique dans la situation actuelle et dans la situation alternative, même si le comportement de l'agent est très différent dans ces deux situations. Cette grande différence nous permet de dire que l'agent est responsable de son omission (car il pouvait agir), alors que la différence insignifiante entre les *résultats* correspondants nous empêche de considérer l'agent comme la cause du résultat actuel (son omission n'expliquant pas ce résultat).

Il est intéressant de remarquer que pour la *causalité déficiente naturelle*, on ne procède pas à une telle comparaison entre la situation actuelle et une autre situation possible. Par exemple, on place la flamme d'une bougie sous une cloche en verre, et elle s'éteint à cause du manque d'oxygène. Supposons que si elle ne s'était pas éteinte de cette manière, elle se serait éteinte une demie seconde après, sous l'effet d'un jet d'eau. Peut-on nier que le manque d'oxygène soit la cause de l'extinction ? Non, même si l'extinction se serait de toute façon produite une demie seconde après. Cette différence entre la causalité du comportement humain et la causalité naturelle s'explique par le fait que les processus naturels, contrairement aux êtres humains, n'ont pas le pouvoir de se retenir d'agir, et ne sont pas responsable de ce qu'ils « font ». Si l'oxygène situé autour de la cloche en verre ne pénètre pas jusqu'à la flamme, c'est qu'il n'est pas responsable de cela, n'ayant pas le pouvoir de le faire. Ainsi, le mot « causer » prendrait un sens particulier lorsque la cause est un agent : un agent S cause E *en tant qu'agent* seulement s'il est responsable de E. Et il cause E à la manière d'une cause naturelle, s'il n'a pas la possibilité de ne pas le causer.

Résumé. En 5.1 et 5.2, nous avons vu que la responsabilité dans les omissions suppose à première vue un pouvoir des alternatives, et qu'elle s'étend aussi loin que ce pouvoir. Quelque soit E (un événement extérieur à S, une omission, ou même un choix de S), S est responsable de E par omission seulement s'il avait le pouvoir d'agir de manière efficace contre E. Si l'agent pouvait seulement choisir autrement, sans faire une action ou causer un événement qui n'existe pas dans le monde actuel, il est seulement responsable de son choix. Là où l'agent n'a pas d'alternative, il n'est pas responsable. Le cas des omissions confirme

ainsi, d'une certaine manière, le principe des possibilités alternatives que j'ai défendu au chapitre II. Plus précisément, il soutient le principe :

(PPA - principe d'action possible) S est responsable de E par omission si E s'est produit, et si S avait la possibilité d'agir efficacement contre E.

Dans les sections 5.3 à 5.5, j'ai abordé trois questions relatives à la nature des omissions. J'ai soutenu d'abord (en 5.3) que la différence entre les actions et les omissions ne se situe pas seulement au niveau de la description des comportements. Un même comportement ne peut être décrit à la fois comme une action et comme une omission, les conditions de la responsabilité morale n'étant pas les mêmes dans les deux cas. Mais la différence de nature entre les actions et les omissions est sans rapport avec le degré d'activité de l'agent. L'action ne suppose pas toujours une activité intense, et inversement : l'omission à l'égard d'un événement peut consister à faire une action très occupante mais inefficace contre cet événement (bricoler au lieu de secourir quelqu'un, organiser une soirée au lieu de préparer un examen). Finalement, la notion d'omission est relative, car ce qui est une omission à l'égard d'un événement peut être une action à l'égard d'un autre événement :

(DAO) (différence entre l'action et l'omission) :

- Un comportement de S qui provoque ou facilite l'événement E est une *action* à l'égard de E.
- Un comportement de S qui ne provoque ni ne facilite l'événement E, mais qui ne s'y oppose pas non plus, est une *omission* à l'égard de E.

En 5.4 et en 5.5, j'ai traité deux questions importantes soulevées par Carolina Sartorio. Tout d'abord, la responsabilité à l'égard d'une omission se transfère-t-elle à ses conséquences de manière causale ? Comme Sartorio, je pense que l'omission peut causer quelque chose, bien qu'elle ne produise pas ses effets par le biais d'un processus physique. Mais contrairement à elle, je pense que la causalité dont il s'agit ici a un aspect contrefactuel : je cause E par omission seulement si E dépend contrefactuellement de mon omission. Autrement dit, mon omission est considérée comme la cause de E parce que si je n'avais pas fait cette omission, E ne se serait pas produit. L'exemple « Aiguillage » utilisé par Sartorio pour

montrer que la responsabilité ne peut pas se transmettre par une relation de dépendance contrefactuelle n'est pas concluant. En l'interprétant comme une action, Sartorio viole son propre critère de distinction entre les actions et les omissions. En plaçant l'aiguillage vers la voie B, l'agent *n'augmente pas* la menace qui pèse sur l'individu ligoté, puisque le train se dirige déjà vers lui. Selon le critère de Sartorio (qui est aussi le mien), l'agent ne fait donc pas une action à l'égard du résultat (la mort de la victime), mais une omission : il n'a pas tourné l'aiguillage vers la voie C, seul moyen d'épargner la victime. Or souvenons-nous que dans « aiguillage », l'agent ne peut pas faire cette action, et que dans une omission, l'agent n'est pas responsable s'il ne peut pas agir de manière efficace. Sartorio devrait conclure qu'il n'est pas responsable de son omission.

Ainsi compris, l'exemple ne contredit pas TR contrefactuel, car :

- La mort de l'individu ligoté sur les rails dépend contrefactuellement de l'omission (consistant à tourner l'aiguillage vers B au lieu de C).
- L'agent n'est pas responsable de la mort de l'individu ligoté sur les rails.

Et contrairement à ce que soutient Sartorio :

- L'agent n'est pas responsable de son omission (à cause de l'intervenant contrefactuel).

Pour que l'exemple contredise TR contrefactuel, il faudrait que l'agent soit responsable de son omission sans être responsable du résultat, ce qui n'est pas le cas. Il est donc possible de soutenir que la responsabilité de l'agent se transfère de l'omission à ses conséquences par le biais d'une relation de dépendance contrefactuelle. En 5.5, j'ai montré que cette idée permettait de résoudre un problème lié à des cas d'omissions où il y a surdétermination causale. Devant ces exemples, il faut renoncer soit au principe de transitivité de la causalité (selon lequel si x cause y et si y cause z, x cause z) soit à TR causal, selon lequel la responsabilité se transmet par le biais d'une relation de causalité. Je pense que la solution la plus satisfaisante est de soutenir que le principe de transitivité de la causalité ne s'applique pas aux omissions parce qu'elles ne causent leurs effets qu'en un sens très particulier. Le

caractère contrefactuel de la relation causale entre l'omission et ses effets explique pourquoi cette relation peut être une exception au principe de transitivité de la causalité.

La théorie ainsi élaborée dans les sections 5.1 à 5.5 va maintenant me servir de base pour examiner les différentes réponses compatibilistes à l'argument des omissions. La première réponse est de réduire les omissions à des actions. Je soutiendrai qu'il est impossible de réaliser cette réduction, notamment à cause des asymétries entre les cas d'actions et les cas d'omissions. Une autre solution est d'accepter le principe PPA, mais en définissant le pouvoir des alternatives de manière conditionnelle, ce qui n'exclut pas le déterministe. La troisième solution est de définir ce pouvoir en termes de mondes possibles. Cette troisième idée sera l'objet du chapitre VI.

5.6 L'approche réductionniste

Nous avons vu que la responsabilité dans les omissions semble liée à la possibilité d'agir autrement. En effet, si un agent S est responsable d'un événement E par omission, c'est parce qu'il avait la possibilité d'agir pour empêcher E. Dans les exemples comme « Requins », où il n'y a pas cette possibilité d'agir, l'agent ne semble pas responsable de E (la noyade, en l'occurrence). Ce constat amène naturellement au principe d'action possible (PPA), qui suppose à première vue l'incompatibilité du libre arbitre et du déterminisme.

Les compatibilistes ont deux possibilités pour résister à cette conclusion. La première est de réduire les omissions à des actions dans tous les cas où l'agent est moralement responsable. La deuxième solution est d'interpréter PPA de manière compatibiliste, en montrant que la possibilité d'agir supposée par PPA peut être définie de telle sorte qu'elle n'exclut pas le déterminisme.

La première de ces deux solutions peut être qualifiée de réductionniste, puisqu'elle réduit les omissions aux actions. On peut la défendre en distinguant plusieurs sortes d'omissions et en montrant ensuite que dans tous les cas où l'agent est dit responsable « par

omission » d'un certain résultat, son omission consiste en une action ou découle d'une action accomplie précédemment par l'agent.

Dans certains cas, on omet de faire A en se retenant d'agir. Ce comportement n'est pas purement passif car il suppose une certaine attention ou concentration. On peut même omettre de faire A en accomplissant volontairement une action qui empêche de faire A. Par exemple :

[Fugitif] Paul est au volant de sa voiture. Il dépasse nettement les limitations de vitesse. En traversant un lieu-dit à 100 km/h, il est surpris par un piéton qui traverse la route. Il freine, mais n'arrive pas à l'éviter. Après le choc, il ralentit un peu et aperçoit dans le rétroviseur le corps du piéton allongé sur la route. Il hésite un moment, puis écrase l'accélérateur pour s'éloigner le plus vite possible.

Paul omet de s'arrêter, et de secourir la victime. Or cette omission est active, car non seulement il ne s'arrête pas, mais il accélère. Il est donc tentant de considérer cette omission comme une action.

Lorsque l'omission ne consiste pas en une action, elle peut souvent être vue comme la conséquence d'une décision ou d'une action volontaire, comme dans l'exemple suivant:

[Anniversaire de mariage] Le jour de ses 25 ans de mariage, Charles omet d'offrir un cadeau à Carole, son épouse. Il a oublié que c'était leur anniversaire de mariage. Or, il y a deux semaines, il a pensé à cet événement, mais après un moment d'hésitation, il a décidé de ne rien noter dans son agenda.¹⁵

Charles est responsable de son omission. Il ne s'est pas retenu d'offrir un cadeau à son épouse. Il n'a pas non plus *décidé* de ne rien lui offrir. Mais il est responsable de son omission parce qu'elle découle d'un de ses choix de manière prévisible. En effet, lorsqu'il a hésité à utiliser son agenda, il a choisi librement de ne pas le faire. Supposons qu'il ait tendance à oublier facilement les dates et les délais et qu'il soit conscient d'avoir ce défaut. Dans ce cas, il est bien responsable de son omission, car il savait qu'en n'utilisant pas son agenda, il risquait

¹⁵ J'emprunte cet exemple à Michael J. Murray, « Pure Omissions, Responsibility, and Character », p. 3.

fort d'oublier l'événement. Nous pouvons dire que l'omission de Charles consistant à ne pas offrir de cadeau à son épouse est une omission non-réfléchie découlant d'une omission réfléchie. La responsabilité de Charles dans la première semble entièrement dérivée de sa responsabilité dans la deuxième, qui lui est antérieure, et peut s'expliquer ultimement par le fait que cette omission précédente a été choisie. Ainsi, nous pouvons distinguer les omissions réfléchies (OR) et les omissions non-réfléchies (ONR) :

(OR) Une omission consistant à ne pas faire A est une omission réfléchie de S si et seulement si S, avant de faire cette omission, s'est représenté l'action A et a choisi de ne pas l'accomplir.

(ONR) Une omission consistant à ne pas faire A est une omission non-réfléchie de S si et seulement si S l'a faite sans s'être représenté l'action A, et par conséquent, sans choisir de ne pas l'accomplir.

La question qui se pose au réductionniste est la suivante : la responsabilité dans les omissions non-réfléchies dérive-t-elle toujours de la responsabilité à l'égard d'une action réfléchie ? Si c'est le cas, le réductionniste pourra dire que les ONR découlent toujours, ultimement, d'un acte de décision, sauf dans le cas d'un oubli ou d'une omission par ignorance dont l'agent ne peut être considéré comme responsable. Ainsi, dans tous les cas où l'agent est responsable d'une ONR, la responsabilité viendrait ultimement d'une action : un acte de décision.

Cette hypothèse peut s'appuyer sur le principe suivant :

(PCC) (principe du choix conscient) On ne peut pas choisir de faire A (ou de ne pas faire A) si on ne se représente pas A.

Imaginons par exemple que dans « Anniversaire de mariage », Charles n'ait jamais pensé, avant la date de ses 25 ans de mariage, que cet événement approchait et qu'il pourrait, à cette occasion, offrir un cadeau à son épouse. Si cela ne lui a jamais traversé l'esprit, comment aurait-il pu penser à offrir ce cadeau ? Il n'avait pas le pouvoir d'y penser. Dans ce cas, il n'est pas responsable de son omission. PCC semble également plausible lorsqu'on l'applique aux

actions. Si je fais quelque chose sans me représenter ce que je fais, cette action peut être volontaire si elle découle d'un choix que j'ai fait avant de l'accomplir. Par exemple, je peux marcher volontairement vers le nord sans y penser : ayant vu sur ma carte que je devais marcher vers le nord pour arriver au bon endroit, je décide d'aller vers le nord, puis je me laisse guider par mon sens de l'orientation. La position du soleil m'indique où se trouve le nord, et si mon sens de l'orientation est suffisamment bon, cet indice constamment présent me permet de marcher vers le nord sans y penser. Mais si j'accomplis cette action sans l'avoir envisagée mentalement à aucun moment, je ne puis en être responsable car ce n'est pas une action volontaire.

En fait l'argument est contestable, même si PCC est vrai, car il suppose une autre idée beaucoup moins évidente :

(PVC) (principe du vouloir conscient) On ne peut faire volontairement A si on ne *choisit* pas de faire A.

En d'autres termes, un comportement volontaire est un comportement choisi consciemment par l'agent. Si PVC et PCC sont vrais, il s'ensuit qu'une omission non réfléchie est toujours une omission dont l'auteur n'est pas responsable, car puisqu'il ne s'est jamais représenté l'action correspondante, il ne pouvait pas *vouloir l'accomplir*.

Or en fait, PVC paraît faux lorsqu'il est confronté à certains genres d'actions aussi bien que d'omissions. Dans une conversation, par exemple, on peut élever le ton volontairement dans le but d'intimider son interlocuteur, et cela sans l'avoir choisi explicitement. Ici, la responsabilité ne découle pas d'un choix délibéré ou pleinement conscient, et il existe des omissions où le même constat s'impose, comme celui-ci, proposé par Murray :

[Randonnée] Deux philosophes, Dan et Mel, ont décidé de se promener en forêt. Après avoir garé leur voiture, ils quittent le parking et s'engagent sur un sentier de randonnée. A ce moment, leur conversation tourne au débat et ils ne voient pas qu'il y a un panneau où il est écrit « Danger ! Serpents venimeux. Faites attention et n'approchez aucun serpent ». Plus loin, Dan marche sur un serpent et se fait mordre. Il meurt un peu plus tard à l'hôpital. Son épouse y rencontre Mel et lui demande ce qui s'est passé. Après

l'explication, elle dit à Mel : « Le gardien du parc assure que le panneau était bien visible. Pourquoi ne l'avez-vous pas vu ? »¹⁶

Mel ne peut pas s'excuser en disant : « en fait, nous l'avons vu – vous parlez d'un panneau écrit en grosses lettres jaunes ? -, nous n'y avons pas fait attention. C'est pourquoi nous ne l'avons pas lu et n'avons pas été sur nos gardes ». ¹⁷

Dans cet exemple, la mort de Dan est la conséquence du fait que Dan et Mel n'ont pas lu le panneau, et cette omission vient d'une certaine négligence. Les deux philosophes paraissent bien responsables du malheur, même s'ils n'ont pas vraiment choisi de ne pas lire le panneau. Ils ont seulement eu le tort de ne pas s'y intéresser. Un panneau comme celui-ci pouvait délivrer des informations importantes. Il fallait donc le lire.

On voit ici ce qu'a de spécifique l'omission par négligence. La négligence ne consiste pas à choisir de ne pas faire attention à quelque chose, mais tout simplement à ne pas y faire attention alors qu'on peut et qu'on doit y faire attention. Le fait que Dan et Mel aient le devoir de faire attention au panneau suppose qu'ils en aient le pouvoir. La conséquence forte tirée du principe PCC est donc fausse. Il n'est pas vrai que :

(conséquence forte) On ne peut pas choisir de faire A si on ne se représente pas A.

La vraie conséquence de PCC est plutôt :

(conséquence faible) On ne peut pas choisir de faire A si on ne *peut pas* se représenter A.

Même si Dan et Mel ne se sont pas représenté clairement l'action de lire le panneau, ils sont responsables de leur omission car ils avaient le pouvoir de se représenter cette action et de la faire. C'est aussi cela qui rend leur omission volontaire. PVC est donc également faux et doit être remplacé par :

¹⁶ Michael Murray & Sam Ruchlewicz. "Without a Trace". Inédit.

¹⁷ *Ibid.*

(PVA) (principe du vouloir alternatif) On ne peut pas faire A volontairement si on ne peut pas vouloir faire non-A.

PVA implique que l'omission de Dan et Mel est volontaire, car ils avaient le pouvoir de vouloir lire le panneau. Mais sans doute ne l'auraient-ils pas s'ils n'avaient pas vu le panneau. Voyant le panneau, ils avaient le pouvoir d'y prêter attention et de le lire. Cela suppose qu'ils avaient une conscience vague mais suffisante du fait qu'il y avait un panneau devant eux, qu'ils pouvaient le lire, et que ce panneau pouvait donner des informations importantes. En somme, le *choix* de faire non-A (c'est-à-dire de ne pas faire A) suppose une représentation claire de non-A, mais pour *faire non-A volontairement*, il suffit d'avoir eu une vague conscience de ce qu'est A.

Comme le note Murray, la responsabilité à l'égard d'une omission *bonne*¹⁸ est possible elle aussi en l'absence d'une représentation claire de l'action correspondante. Le manque d'attention à un certain objet ou à une certaine action possible peut être méritoire, comme dans cet exemple :

[Betty patriote] Betty naît et grandit dans un milieu marqué par un fort patriotisme. Comme la plupart des gens, elle intériorise, au cours de son éducation, les valeurs de son milieu. A aucun moment, elle n'a fait le choix d'être fidèle aux valeurs patriotiques de son milieu, mais elle aurait pu les rejeter bien des fois. Par exemple, lorsqu'elle avait quinze ans, elle a vu un documentaire sur la transmission des valeurs nazies aux jeunes allemands pendant les années 30. Elle aurait pu en conclure qu'elle devait rejeter les valeurs de son milieu. Mais elle ne l'a pas fait. Un jour, un homme lui propose une forte somme d'argent pour l'inciter à trahir son pays, mais elle refuse cette proposition.¹⁹

¹⁸ J'ai supposé plus haut qu'on ne peut jamais être responsable d'un événement par omission si cet événement est bon. Cela n'empêche pas de croire qu'on peut être responsable de l'omission bonne elle-même.

¹⁹ Je m'inspire ici d'un exemple de Fischer et Ravizza. Cf. J. M. Fischer & M. Ravizza, "When the Will is Free", *Philosophical Perspectives*, ed. J. Tomberlin., vol.6, Ridgview Publishing, Atascadero, CA, p. 423-451. Repris dans O'Connor (1995a). L'exemple se trouve p. 260 dans l'anthologie de O'Connor.

Supposons que Betty n'ait pas été manipulée par son milieu et que son choix ne soit pas déterminé causalement. Betty semble alors responsable de son refus de trahir. Cela s'explique par le fait qu'elle aurait pu accorder plus d'intérêt à l'argent et en accorder moins à sa patrie. Si l'on n'admet pas ce genre d'explication, on peut difficilement rendre compte de la formation des vices et des vertus, car si la responsabilité d'un agent à l'égard de ses dispositions supposait qu'il les ait choisies, il n'en serait que très peu responsable. En effet, nous acquérons une bonne part de nos dispositions sans les choisir délibérément mais en accordant plus d'intérêt à certaines choses qu'à d'autres. Les exemples « Randonnée » et « Betty » montrent que certaines omissions non-réfléchies sont irréductibles à des actions, même si nous comptons parmi les actions le simple fait de prendre une décision. La responsabilité dans de telles omissions ne suppose pas une *décision* de l'agent, et au fond, cela n'est pas surprenant car il y a aussi des *actions* dont nous sommes responsables sans avoir décidé de les accomplir. En dernier ressort, la responsabilité à l'égard d'omissions et d'actions que nous n'avons pas décidé de faire s'explique par un certain contrôle que nous possédons sur ces actions ou omissions.

5.7 Le problème des asymétries et l'argument de l'exclusion

Nous avons vu que certaines omissions dont l'auteur est responsable ne viennent pas d'un acte préalable de décision. Elles relèvent de la simple négligence ou inattention. Notons d'abord que si tous les autres genres d'omissions pouvaient se réduire à des actions, cela poserait de grandes difficultés théoriques et pratiques. En effet, il y a des asymétries entre les cas d'actions et les cas d'omissions dans nos jugements intuitifs sur la responsabilité morale. Or si toutes les omissions dont l'auteur est responsable se ramenaient en dernière analyse à des actions, ces asymétries ne devraient pas exister. En particulier, on s'accorde à reconnaître que dans les exemples d'actions à la Frankfurt, la responsabilité de l'agent à l'égard du résultat de son action ne suppose pas qu'il ait eu la possibilité d'éviter ce résultat, compris comme un universel²⁰. Au contraire, dans les cas d'omissions du même type que « Requins », on

²⁰ Cf. l'exemple de Van Inwagen discuté en 2.2 : Gunnar tire sur Ridley. Ridley meurt. Gunnar n'aurait pas pu ne pas tirer, car un savant fou contrôlait ses choix et ses actions avec l'intention qu'il tue Ridley. Gunnar tue Ridley

s'accorde à dire que l'agent n'est pas responsable du résultat (par exemple la noyade) parce qu'il ne pouvait pas l'empêcher de se produire.

Plus haut, j'ai montré comment il faut comprendre cette asymétrie face à certains cas frankfurtiens d'omissions dans lesquels l'agent pourrait sembler responsable alors même qu'il ne peut rien faire d'efficace pour empêcher le résultat²¹. Dans « Deux tueurs 1 », où un individu omet de protéger Ridley contre Gunnar, mais où un autre tireur l'aurait abattu si Gunnar n'avait pas tiré, notre jugement sur la responsabilité de l'agent dépend du degré de différence entre le résultat actuel et le résultat qui serait advenu s'il avait empêché le résultat actuel. Si la différence entre ces deux résultats est négligeable, la responsabilité de l'agent l'est aussi, car il ne pouvait pas éviter que se produise un résultat à peu près identique au résultat actuel.

La thèse réductionniste se fonde sur le fait que beaucoup d'omissions :

- (a) consistent à se retenir d'agir, ce qui serait une action. Ou :
- (b) consistent en une action (par exemple, en faisant la fête, un étudiant peut omettre de préparer ses examens). Ou :
- (c) découlent d'une action volontaire (par exemple, si à t₁, on se rend volontairement incapable d'accomplir une certaine action à t₂, comme quelqu'un qui boit avant de conduire, et se rend ainsi incapable de freiner à temps).
- (d) découlent d'un acte de décision (comme celle de Jean dans « Requins »).

Or si nous considérons des exemples assez simples où l'omission appartient à l'une de ces catégories, nous pouvons remarquer que cette modification ne change pas notre jugement sur la responsabilité de l'agent. Dans tous les cas, l'asymétrie subsiste par rapport aux exemples que tous considèrent comme des exemples d'actions. Ainsi, dans « Requins », on

par lui-même, sans que le savant fou ait besoin d'intervenir. Il est visiblement responsable de son action et de la mort de Ridley, même s'il n'avait pas la possibilité d'empêcher que Ridley meure assassiné par lui. En revanche, on peut considérer avec Van Inwagen que l'événement « Ridley meurt » (ou « Ridley meurt assassiné à t ») n'est pas le même dans les deux cas, car il n'a pas la même histoire causale. Gunnar a donc une possibilité alternative : celle de tuer Ridley *sous l'influence de l'intervenant contrefactuel*. En faisant cela, Gunnar aurait pu éviter d'engager sa propre responsabilité dans la mort de Ridley.

²¹ Cf. *supra*, p. 223 et s.

peut imaginer que Jean se retient d'agir ou s'éloigne en courant pour ne pas sauver l'enfant, mais cela ne le rend pas responsable de la noyade. En plus d'être responsable d'une mauvaise intention et d'une omission (parce qu'il n'essaie pas de sauver l'enfant), il est responsable de courir ou de se retenir d'agir. Cela le rend peut-être encore plus blâmable, mais dans un cas comme dans l'autre, il n'est pas plus responsable de la noyade que s'il faisait une simple omission.

Les omissions du type (c) constituent une exception car dans ce genre de cas, l'agent est responsable de manière dérivée. C'est pourquoi même s'il ne peut agir de manière efficace, il est responsable de son omission et du résultat. Voici un cas de ce genre :

[Promenade des requins] Jean ouvre l'enclos des requins pour qu'ils profitent d'un espace plus grand pendant quelque heures. Le lendemain, il voit un enfant sur le point de se noyer. Il veut le sauver, mais c'est impossible, car il a oublié de faire rentrer les requins dans l'enclos.

Jean paraît bien responsable de son omission et de la noyade. Cela s'explique par le fait qu'il est responsable de sa propre incapacité d'agir. Donc si ce cas ne présente pas d'asymétrie par rapport aux actions, c'est probablement parce qu'il s'agit ultimement d'une action entraînant une incapacité d'agir à un moment ultérieur. En revanche, pour les omissions non-réfléchies et toutes les omissions de type (a), (b), et (d), l'asymétrie subsiste.

Ce qui pourrait jeter le doute sur ces asymétries, c'est l'impression qu'il n'y a qu'une distinction de degré entre l'action et l'omission par inattention. Pour le réductionniste, les omissions et les actions appartiendraient à un continuum d'activité que l'on peut schématiser ainsi :

- action
- omission consistant en une action
- omission consistant à se retenir d'agir
- omission découlant d'un acte de décision
- omission par inattention découlant d'une action antérieure ou d'une décision antérieure

Les omissions « pures » ne résultant pas d'une décision de l'agent existeraient mais ne seraient pas imputables à l'agent. Elles viendraient d'oublis ou d'un manque de connaissance dont l'agent n'est pas responsable.

En réalité, cette hypothèse ne tient pas, même si les omissions impliquent un certain degré d'activité dans le comportement de l'agent. En effet, comme nous l'avons vu plus haut, il y a une discontinuité claire entre les comportements qui favorisent un certain effet et ceux qui ne s'opposent pas à cet effet. Par exemple, il y a une discontinuité entre le cas où je tue Ridley, et celui où je laisse Gunnar le tuer, même si j'accomplis une action au moment où je laisse Gunnar le tuer. Si je fais des mots croisés au moment où Gunnar tire sur Ridley, il reste vrai que mon action de faire des mots croisés ne provoque ni ne facilite la mort de Ridley. Il s'agit donc d'une omission à l'égard de cet événement. Mais si mon action de faire des mots croisés fait partie du plan de Gunnar pour tuer Ridley et qu'en faisant cette action, je participe volontairement à l'assassinat, alors je suis responsable *par mon action* de la mort de Ridley. Ainsi, comme je l'ai dit plus haut, le fait qu'un comportement soit une action ou une omission à l'égard d'un certain événement E n'est pas fonction du degré d'activité de l'agent dans ce comportement, mais du fait que ce comportement provoque (ou facilite) E, ou ne s'oppose pas à E. S'il ne provoque pas E mais ne s'y oppose pas non plus, il s'agit d'une omission. En un sens, une omission peut *contribuer* à un événement (c'est pourquoi on peut être *responsable* d'un événement par omission), mais cette contribution s'analyse de manière contrefactuelle. Il y a donc bien une différence nette entre les actions et les omissions.

Carolina Sartorio²² a formulé une autre objection contre la thèse réductionniste en s'inspirant du problème de l'exclusion du mental, problème que Jaegwon Kim a mis en évidence pour critiquer le matérialisme émergentiste²³. De ce modèle, nous pouvons tirer deux objections inverses. Je commencerai par celle que Sartorio n'a pas formulée, mais en reprenant plusieurs de ses remarques sur la spécificité des omissions.

²² Carolina Sartorio, "Omissions and Causalism", *Noûs* 43/3 (2009), p. 513-530.

²³ Cf. Jaegwon Kim, *Trois essais sur l'émergence*, trad. Mathieu Mulcey, Ithaque, Paris, 2005.

Tout comme la réduction des phénomènes mentaux aux phénomènes physiques, la réduction des omissions aux actions se heurte à un problème d'exclusion. Certains matérialistes pensent que les événements mentaux dépendent entièrement des événements physiques dans leur existence, et que tous les comportements humains faisant intervenir l'esprit peuvent être expliqués en termes purement physiques, mais sont pourtant irréductibles aux événements physiques, et « émergent » sur eux. Kim a montré que cette position est difficilement tenable car si les phénomènes mentaux ont une explication en termes purement physiques, cette explication suffit, et l'explication en termes mentaux ne renvoie à aucune réalité autonome par rapport aux réalités physiques. Un problème du même genre se pose si on admet les deux affirmations suivantes :

- les omissions ont une spécificité par rapport aux actions. En particulier, elles n'ont pas les mêmes pouvoirs causaux.
- Les omissions peuvent être décrites comme des actions.

Ces affirmations sont incompatibles, car si les omissions peuvent être décrites correctement comme des actions, elles ne renvoient à aucune réalité distincte, et la première affirmation est fausse. Or la première affirmation semble vraie pour plusieurs raisons. Premièrement, nous avons vu que nos jugements sur la responsabilité impliquent des asymétries entre les actions et les omissions. Deuxièmement, supposons que je sois à la place de Jean dans « Noyade ». Si j'omets de sauter à l'eau en mangeant une glace (donc en faisant une autre action à la place), cela ne veut pas dire que le fait que je mange une glace soit cause du fait que je ne saute pas à l'eau. J'aurais d'ailleurs pu faire cette omission d'une autre manière : en lisant un livre, en faisant des pompes, etc. Troisièmement, actions et omissions ont des pouvoirs causaux différents. Par exemple, le fait que je ne saute pas à l'eau ne peut pas causer des maux d'estomac le lendemain, contrairement au fait que je mange une glace. Quatrièmement, il y a des cas où aucune action ne peut être identifiée à l'omission : ce sont les omissions *par inattention*. Cinquièmement, contrairement aux actions, les omissions causent leurs effets de manière contrefactuelle.²⁴

²⁴ Cf. *supra*, 5.4 (p. 226 et s.).

En tenant compte de toutes ces différences, on peut construire un argument montrant que les réductionnistes ne peuvent pas rendre compte de la spécificité des omissions. Mais on peut aussi raisonner de manière inverse, comme le fait Sartorio.²⁵ L'argument montre alors que puisque les omissions peuvent causer des événements, il n'est pas nécessaire de faire référence à des actions pour expliquer la responsabilité d'un individu à l'égard d'un événement. Considérons l'omission :

O2 : Jean omet de sauter à l'eau.

Cette omission est précédée de :

O1 : Jean omet de former l'intention de sauter à l'eau.

Pour le réductionniste, il y a une action A telle que :

(Thèse 1) A a causé O2.

Mais on peut tout aussi bien penser que :

(Thèse 2) O1 a causé O2.

Qu'est-ce qui cause O2, puis la mort de l'enfant ? Une omission mentale, ou une action mentale ? Selon Sartorio, la thèse 2 est vraie et sa vérité sape les fondements de la thèse 1. Cette dernière idée est la thèse de « l'exclusion causale appliquée aux omissions » (ECAO).

Pour établir ECAO, on considère d'abord ces deux assertions :

(Thèse 3) A a causé la mort de l'enfant.

(Thèse 4) O2 a causé la mort de l'enfant.

La première prémisse de l'argument est :

(P1) La thèse 4 est vraie et sa vérité exclut la vérité de la thèse 3.

²⁵ Je reprends ici l'argument de Carolina Sartorio dans "Omissions and Causalism", p. 517 et s.

Ensuite, Sartorio raisonne ainsi²⁶ :

Si la vérité de la thèse 4 est suffisante pour que la thèse 3 soit fausse, alors, de même, la vérité de la thèse 2 est suffisante pour réfuter la thèse 1. La vérité de la thèse 4 montre que les omissions peuvent être des causes. Et si les omissions peuvent être des causes, alors l'explication proposée par la thèse 2 est suffisante et il n'y a pas besoin de recourir à l'explication proposée par la thèse 1. Donc il n'y a pas d'action qui soit la cause de O2.

Par exemple, considérons cette action : « Jean forme l'intention de ne pas sauter à l'eau ». Elle n'est pas la cause de O2. En effet, la cause de O2 n'est pas le fait que Jean *ait eu* l'intention de *ne pas* sauter à l'eau, mais le fait que Jean *n'ait pas eu* l'intention de sauter à l'eau.

La deuxième prémisse du raisonnement est :

(P2) Si P1, alors la thèse 2 est vraie et sa vérité exclut la vérité de la thèse 1.

De P1 et P2, on déduit enfin ECAO.

Tout l'argument de Sartorio repose sur la thèse 4, c'est-à-dire sur l'idée que dans certains cas, une omission suffit à expliquer un certain événement. Autrement dit, une omission peut être la cause d'un événement, et il est donc superflu de postuler une action mentale comme cause de l'événement. Le principe du rasoir d'Occam interdit de postuler des entités sans nécessité ; il exclut donc ici de ramener l'événement ou l'omission à une action.

A cette objection, le causaliste peut répondre ainsi :

- l'enfant s'est noyé parce que *Jean n'a pas sauté à l'eau* (O2).
- *Jean n'a pas sauté à l'eau* (O2) parce que *Jean mangeait une glace* (A).

Donc il semble qu'on puisse conclure, par transitivité :

- L'enfant s'est noyé parce que *Jean mangeait une glace* (A).

²⁶ *Ibid.*, p. 519.

Mais ce raisonnement ne tient pas. En effet, A implique O2 mais ne cause pas O2 pour autant. Donc l'argument ne prouve pas que O2 cause la noyade de l'enfant. De plus, A et O2 peuvent être exactement simultanés. Cela suggère que A ne cause pas O2, car une cause *précède* son effet. Le causaliste peut-il dire que A cause l'événement *directement*, sans causer d'abord O2 ? Cette idée est peu plausible, comme on le voit ici :

[Révisions] Jim a invité ses amis pour une soirée festive. Par conséquent, il n'étudie pas en vue de son examen. Le lendemain, il arrive à l'heure pour son examen, mais n'a pas les connaissances requises. Deux ou trois heures de travail lui auraient suffi pour obtenir une note passable. Mais comme il n'a pas révisé la veille au soir, il échoue lamentablement.

Est-ce l'action de Jim de faire la fête qui cause son échec aux examens ? Non, la cause immédiate, et qui suffit, est le fait qu'il ne révise pas. Elle pourrait suffire car Jim peut ne pas réviser sans faire la fête, tout comme il peut faire cette omission d'une autre manière : en lisant des romans ou en regardant la télévision. La cause de son omission ou de son échec aux examens n'est donc pas l'action de faire la fête, mais ce qu'*implique* cette action. Ainsi, la relation entre « Jim fait la fête à t » et « Jim n'étudie pas à t » n'est pas une relation de cause à effet, mais une relation d'implication.

Cet argument de l'exclusion repose finalement sur le fait qu'il existe des omissions qui ne s'accompagnent pas d'une action et qui ne sont pas l'effet d'une action : ce sont les omissions *par inattention*, dont il a été question dans la section précédente. Ces omissions ne découlent même pas d'un acte de décision, contrairement à d'autres omissions non-réfléchies. Si les omissions pas pure inattention existent, il s'ensuit que les omissions dont on ne voit pas la cause n'ont pas besoin d'être expliquées à partir d'une action. Cela ne signifie pas qu'aucune omission ne découle d'une action, mais que l'explication réductionniste est superflue. Il faut donc admettre que beaucoup d'omissions sont des omissions « pures », irréductibles aux actions, et cela nous ramène à notre hypothèse de départ, à savoir que la responsabilité dans les cas d'omissions suppose un pouvoir des alternatives.

Nous avons vu également que si la thèse réductionniste était vraie, les conditions de la responsabilité morale devraient être les mêmes pour les actions et les omissions, ou du moins, un même cas pourrait être considéré comme une action ou comme une omission sans que cela modifie notre jugement sur la responsabilité de l'agent. Or ce n'est pas le cas. Par exemple, nous avons vu que dans « Princesse », notre jugement sur la responsabilité de la princesse n'est pas le même s'il s'agit d'une action ou s'il s'agit d'une omission : dans le premier cas, elle est responsable, et non dans le deuxième.²⁷

Ces deux objections prouvent qu'il faut considérer les omissions comme un phénomène autonome et irréductible aux actions. Compte tenu de cette spécificité, les compatibilistes ont le choix entre deux solutions pour éviter une compréhension libertarienne du principe PPA. La première est de dire que la possibilité d'agir dont dispose un agent lorsqu'il est responsable d'un événement par omission se comprend de manière conditionnelle. La deuxième consiste à analyser cette possibilité en termes de mondes possibles. Dans la prochaine section, je m'intéresserai à la première de ces deux solutions. J'aborderai la deuxième dans le chapitre suivant en examinant plus particulièrement la thèse de Fischer et Ravizza.

5.8 L'analyse conditionnelle de la responsabilité dans les omissions

Lorsqu'un individu est responsable par omission d'un événement E, sa responsabilité n'est pas liée comme dans le cas d'une action à une séquence causale actuelle allant de son comportement à l'événement E, car il ne fait rien qui provoque ou facilite l'événement E. Tout ce qu'on lui reproche, si l'événement E est malheureux, c'est d'avoir laissé E se produire, ce qui revient à dire qu'on lui reproche de ne pas avoir fait une action en faveur de non-E. La responsabilité dans les omissions a donc un aspect contrefactuel, et le défi pour les théories compatibilistes est de rendre compte de cet aspect contrefactuel sans concéder que la responsabilité dans les omissions suppose, au sens le plus strict, la possibilité d'agir autrement. Cela reviendrait en effet à déclarer l'incompatibilité du déterminisme avec la

²⁷ Cf. *supra*, p. 223 et s.

responsabilité morale dans les omissions. En somme, le compatibilisme doit élaborer sa propre interprétation des conditions suivantes :

S est responsable de E par omission seulement s'il y a une action A telle que :

- (a) S ne fait pas A et E se produit.
- (b) S peut faire A.
- (c) Si S faisait A, E ne se produirait pas.

Ce qui oppose les compatibilistes et les incompatibilistes est l'interprétation de la condition (b). Pour les incompatibilistes, (b) exprime une possibilité alternative *strictement accessible* à l'agent, ce qui exclut la vérité du déterminisme. Pour reformuler (b), les compatibilistes ont le choix entre deux solutions : l'analyse conditionnelle du verbe « pouvoir », et l'analyse en termes de mondes possibles. Dans cette section, j'examinerai seulement la première solution, défendue par Randolph Clarke et Alison McIntyre.

Dans une analyse *conditionnelle* du verbe « pouvoir », la condition (b) se reformule ainsi :

- (b') Si S voulait faire A, il ferait A.

Mais l'ensemble des conditions (a), (b') et (c) ne suffit pas, car il faut y ajouter les conditions épistémiques de la responsabilité morale. Un individu S qui remplit les conditions (a), (b') et (c) n'est pas responsable de E s'il ignore qu'il doit faire A et qu'il peut faire A. Une définition conditionnelle plus complète de la responsabilité dans les omissions doit donc ressembler à ceci :

(RO conditionnel) S est responsable par omission de l'événement E si et seulement si il y a une action A telle que :

- (a) S ne fait pas A et E se produit.
- (b') Si S voulait faire A, il ferait A.
- (c) Si S faisait A, E ne se produirait pas.

(d) S sait qu'il peut faire A et que s'il faisait A, E ne se produirait pas.

(e) S sait qu'il faut empêcher E de se produire.

Les conditions (d) et (e) sont les conditions épistémiques de la responsabilité dans les omissions. RO conditionnel ne diffère ainsi de RO que par (b'), l'analyse conditionnelle de (b). Alison McIntyre, qui défend une interprétation conditionnelle et compatibiliste de (b), a énoncé un ensemble de conditions sensiblement différent de celles que je viens d'énoncer.

Sa première condition est celle-ci :

« (a) L'agent a décidé de ne pas faire A au terme d'un processus ordinaire de délibération ». ²⁸

Cette condition est floue, car McIntyre ne définit pas ce qu'est un « processus ordinaire de délibération ». De plus, cette condition est trop forte, car il existe des omissions par inattention, comme nous l'avons vu plus haut. Dans « Randonnée » ²⁹, Dan et Mel sont responsables par omission de la mort de Dan, n'ayant pas lu un panneau qu'ils avaient le devoir de lire. Or cette omission n'est pas délibérée et ils n'ont même pas décidé de ne pas lire le panneau.

Il faut donc une condition plus faible, exigeant seulement que S fasse l'omission volontairement en ayant toutes les connaissances lui permettant de comprendre qu'il doit faire A pour empêcher E. Ces conditions sont (b), (d), et (e). (d) et (e) indiquent ce qu'il doit savoir pour *vouloir faire A*, et (b) suppose qu'il omet volontairement de faire A, car s'il peut faire A et sait qu'il doit le faire, son omission est volontaire. Mais si l'on remplace (b) par (b'), comme dans la version conditionnelle de RO, il faut préciser dans une condition supplémentaire que S omet *volontairement* de faire A, car cela n'est pas impliqué par (b'), même si on lui adjoint les

²⁸ Alison McIntyre, "Compatibilists Could Have Done Otherwise: Responsibility and Negative Agency". *The Philosophical Review*, Vol.103/3 (1994): 453-488. Le passage que je cite se trouve p. 466.

²⁹ Cf. *supra*, p. 242.

autres conditions énoncées dans RO conditionnel. En effet, la proposition « Si S voulait faire A, il le ferait » est compatible avec « S ne veut pas faire A ». Mais surtout, (b') est insuffisant pour les deux raisons suivantes. Premièrement, avec (b'), l'ensemble des conditions n'exclut pas les cas où l'agent fait son choix sous l'effet d'une drogue, de l'hypnose ou d'une influence aliénante. Par exemple, il peut choisir de ne pas faire A parce qu'il a consommé une drogue qui le rend fainéant et pessimiste. Deuxièmement, dans cette sorte de cas, l'agent ne peut pas décider de faire A, car il n'en est pas capable. Pour ces deux raisons, il n'est pas moralement responsable de son omission. J'ai déjà montré ce défaut de l'analyse conditionnelle au chapitre IV³⁰. Un libertarien complète (b') en précisant que S a la possibilité *au sens le plus strict* de faire A. Un compatibiliste ne peut pas dire cela (sinon il renonce à sa thèse) mais il peut compléter (b') par cette condition :

(b'2) Il y a au moins un monde possible où S veut faire A pour empêcher E de se produire.

Si cette condition était formulée de manière conditionnelle, cela donnerait :

Si S voulait vouloir faire A pour empêcher E, il voudrait faire A.

Or cette formule destinée à compléter (b') tomberait sous la même objection que (b') : dans un cas comme dans l'autre, la vérité du conditionnel n'implique pas la vérité de l'antécédent ; on peut donc imaginer des exemples où l'agent remplit toutes les conditions énoncées, alors qu'il ne peut pas vouloir *faire* A (ou vouloir *vouloir faire* A).³¹ Si l'agent a acquis un désir incontrôlable de faire A, et qu'il l'a acquis involontairement, il n'est pas responsable de faire A, même s'il est vrai que *s'il voulait se retenir de faire A, il ne ferait pas A*. La même objection s'applique aux omissions. Si on m'a forcé à consommer une drogue très puissante qui provoque une paresse irrésistible, on ne peut pas me reprocher de ne pas réviser la grammaire hébraïque après 22h, même s'il est vrai que *si je voulais le faire, je le ferais*. Pour être responsable de cette omission, il faut que je puisse réellement vouloir étudier l'hébreu après 22h.

³⁰ Cf. 4.3, p. 186.

³¹ Cf. l'exemple de Jojo utilisé par Suzan Wolf.

Pour éviter cet inconvénient, il faut formuler compléter la condition (b') en faisant référence à des situations possibles ou « mondes » possibles. C'est ce que fait Alison McIntyre.³²

On obtient ainsi cette définition, où (b') est complétée par (b'2) :

(RO conditionnel modifié) S est responsable par omission de l'événement E si et seulement si il y a une action A telle que :

(a) S ne fait pas A et E se produit.

(b') Si S voulait faire A, il ferait A.

(b'2) Il y a au moins un monde possible où S veut faire A pour empêcher E de se produire.

(c) Si S faisait A, E ne se produirait pas.

(d) S sait qu'il peut faire A et que s'il faisait A, E ne se produirait pas.

(e) S sait qu'il faut empêcher E de se produire.

Cette définition des conditions de la responsabilité dans les omissions présente encore un défaut. Comme dans RO, que j'ai exposé plus haut³³, les conditions épistémiques (d) et (e) sont trop fortes, car on peut être responsable d'un événement par omission sans savoir que cet événement doit être évité, ou sans savoir comment on peut l'éviter. Par exemple, si j'omets d'aller à une réunion professionnelle importante parce que j'ignore que cette réunion était prévue, cette omission peut venir du fait que je n'ai pas pris connaissance de toutes les informations qu'on m'a transmises par e-mail, et dans ce cas, je suis responsable de mon omission parce que même si je ne savais pas que cette réunion était prévue, je *pouvais* le savoir. Il faut donc corriger (d) et (e).

³² La deuxième condition énoncée par McIntyre est la suivante : “In *some* situation in which the agent believed that there was reason to do A, the agent would have decided to do A through a process of ordinary deliberation” (*Ibid*, p. 466). Fischer et Ravizza utilisent un critère du même genre, que j'examinerai dans le chapitre VI.

³³ Cf. *supra*, p. 218.

Dans RO*, je les ai corrigées en remplaçant à chaque fois « sait » par « peut savoir »³⁴. Ici, cette opération est superflue et compliquerait inutilement les formules. Elle les compliquerait inutilement parce que pour éviter encore une fois le sens libertarien du mot « peut », il faudrait remplacer ce mot par une formule conditionnelle, ce qui donnerait :

(d'') Si S voulait savoir comment éviter E, il le saurait.

(e'') Si S voulait savoir si E est à éviter, il le saurait.

La solution la plus simple est de supprimer (d') et (e') en mettant tout le contenu de ces deux conditions dans (b'2). On obtient ainsi :

(RO* conditionnel) S est responsable par omission de l'événement E si et seulement si il y a au moins une action A telle que :

(a) S ne fait pas A et E se produit.

(b') Si S voulait faire A, il ferait A.

(b'2) Il y a au moins un monde possible où S veut faire A pour empêcher E, sachant qu'il faut éviter E et que faire A permet d'éviter E.

(c) Si S faisait A, E ne se produirait pas.

Randolph Clarke et Alison McIntyre défendent une conception de la responsabilité dans les omissions qui est très proche de celle-ci. Chacun d'eux la justifie par une comparaison entre deux exemples. Clarke présente d'abord le cas suivant :

[« Babysitter »] Sam promet de garder le petit Freddy, mais oublie sa promesse. Personne ne provoque cet oubli, mais cette promesse lui sort de la tête. Par conséquent, il omet de se présenter à la porte pour garder le petit Freddy. A l'insu de Sam, un savant fou dirige ses pensées. Si Sam avait tendance à se souvenir de

³⁴ Cf. *supra*, p. 222.

sa promesse, le savant interviendrait et l'empêcherait de s'en souvenir. En l'occurrence, le savant n'intervient pas puisqu'il n'en a pas besoin³⁵.

Selon Clarke, Sam est responsable de l'omission, car s'il ne se présente pas à la porte pour le babysitting, c'est parce qu'il a oublié sa promesse. Or cet oubli n'est pas dû à l'influence d'une cause extérieure ou échappant au contrôle de Sam. Celui-ci est donc responsable de l'oubli et de l'omission qui en résulte. Pourtant, il est clair qu'il ne pouvait pas agir autrement, puisque le savant fou l'en aurait empêché. L'exemple de Clarke paraît alambiqué. Il n'est pas nécessaire, pour établir cette conclusion, d'imaginer que le savant fou empêche Sam de se souvenir de sa promesse. Il suffit qu'il l'empêche de s'en souvenir assez longtemps, ou bien de marcher dans la bonne direction. Mais pour Clarke, la situation ne serait plus la même : Sam ne serait plus responsable de l'omission. Clarke admet en effet le principe suivant :

S est responsable d'une omission (ne pas faire A) → Si S voulait faire A, il ferait A.

Mais cette idée a des implications étranges. Elle suggère en effet que Sam n'est pas responsable de ses actes si le savant fou contrôle ses mouvements sans contrôler ses choix, alors qu'il en est pourtant responsable si le savant fou contrôle à la fois ses mouvements et ses choix. L'analyse conditionnelle s'écarte ainsi de l'intuition selon laquelle plus un individu est soumis à l'influence de causes, moins il est responsable de ses actes. Surtout, elle se heurte aux exemples où un individu est aliéné jusqu'en ses désirs³⁶.

Une position voisine de celle de Clarke a été soutenue par Alison McIntyre. Elle compare deux exemples, que je reprends ici en les simplifiant :

[Pompier 1] Vous êtes un pompier et un important feu de forêt s'approche par le nord. Vous croyez que vous pouvez allumer un autre feu (contrôlé) pour créer une zone où l'incendie de forêt s'arrêtera faute de combustible. Quand les flammes de l'incendie

³⁵ Randolph Clarke, « Ability and Responsibility for Omissions », *Philosophical Studies* 73 (1994), p. 203.

³⁶ Cf. les exemples donnés un peu plus haut : si on me force à consommer une drogue provoquant une paresse irrésistible, je ne suis pas responsable du fait que je ne révise pas la grammaire de l'hébreu après 22h, même s'il est vrai que *si je voulais le faire, je le ferais*.

arriveront à cet endroit, elles ne rencontreront en effet que des plantes calcinées. Pour allumer ce feu, vous voulez utiliser le réservoir à essence de votre camion, mais finalement, vous renoncez à allumer ce feu. Du coup, l'incendie avance toujours vers le sud, détruisant de larges surfaces de forêt. En fait, le réservoir du camion a fui, et il est entièrement vide au moment où vous renoncez à mettre le feu. Vous ne pouvez donc arrêter l'incendie.³⁷

Dans ce cas, comme l'admet McIntyre, vous n'êtes pas responsable du fait que l'incendie progresse, car vous ne pouvez pas allumer de contre-feu. Modifions maintenant l'exemple :

[Pompier 2] Cette fois-ci, le réservoir contient suffisamment d'essence, mais vous décidez de ne pas allumer de feu. Or des écologistes fanatiques, qui luttent sans aucun discernement contre les feux de forêt, ont payé un neurologue surdoué pour contrôler vos délibérations. Vous décidez spontanément de ne pas allumer de feu, donc le neurologue n'intervient pas, mais dans le cas contraire, il aurait repéré une inclination dans votre cerveau et l'aurait arrêtée en déclenchant artificiellement la décision de ne pas allumer de feu. Vous ne pouviez donc pas intervenir.

McIntyre estime que dans cet exemple, vous êtes responsable de l'omission, alors que dans le premier, vous ne l'étiez pas. En effet, vous avez décidé par vous-même de ne pas intervenir, et si vous aviez pris la décision contraire, rien ne vous aurait empêché d'allumer le feu³⁸. C'est aussi ce que dirait Clarke, et l'objection que j'ai faite à propos de son exemple vaut également ici : en fait, vous ne pouviez pas allumer de contre-feu puisque vous ne pouviez même pas *décider* de le faire. Le neurologue vous empêchait de prendre cette décision. Aussi, l'interprétation de l'exemple par McIntyre est contestable. Dans « Pompier 2 », comme dans « Pompier 1 », vous êtes blâmable, mais sans être responsable de l'accident³⁹. Vous êtes sans

³⁷ Alison McIntyre, "Compatibilists Could Have Done Otherwise: Responsibility and Negative Agency", p. 458.

³⁸ *Ibid.*, p. 466.

³⁹ Fischer, Ravizza, et Byrd refusent également de faire une différence entre « Pompier 1 » et « Pompier 2 », ou « Babysitter 1 » et « Babysitter 2 », mais pour une autre raison que moi. Selon eux, il n'y a pas à tenir compte du niveau où se place l'intervenant contrefactuel (avant ou après le moment où l'agent essaie d'agir), car l'agent a le contrôle de direction dans les deux cas, et l'intervenant contrefactuel n'intervient pas dans la séquence actuelle.

doute responsable de votre attitude, de votre paresse, de l'omission consistant à ne pas vouloir sauver l'enfant, et de votre décision de ne pas le sauver, mais vous n'êtes pas responsable du fait que l'incendie progresse vers le sud. Dans « Pompier 2 », le contrôle que vous avez sur votre comportement est encore plus limité que dans « Pompier 1 », puisque vous ne contrôlez même plus vos choix. Il est donc étrange de soutenir que vous êtes responsable de votre omission ou de l'incendie dans « Pompier 2 », alors que vous ne l'êtes pas dans « Pompier 1 ».

Ceci montre que la condition (b') utilisé dans RO* conditionnel (et admise par Clarke et McIntyre) n'est pas satisfaisante. Or RO* pose encore deux problèmes. Premièrement, la formulation de (b'2), complément indispensable de (b'), n'exclut pas les mondes possibles où l'agent décide de faire A pour de bonnes raisons, mais sous l'influence d'une drogue, de l'hypnose, ou d'une autre influence aliénante. Or le fait qu'un tel monde possible existe ne prouve nullement la responsabilité de l'agent. Par exemple, imaginons ceci :

[Syndicat 1] Jean et Pierre font partie du même syndicat. Un nouveau responsable syndical va être élu lors de la prochaine réunion. Jean hésite à voter pour Luc. Finalement, il s'abstient. Un autre candidat beaucoup moins bon est élu, dont les décisions sèment la discorde entre les membres du syndicat. Or il y a un monde possible où Jean vote pour Luc, pour d'excellentes raisons, mais sous l'influence de l'hypnose. En effet, dans ce monde possible, Pierre hypnotise Jean pour le convaincre plus facilement de voter pour Luc.

Visiblement, le fait que Jean vote pour Luc dans ce monde possible ne nous renseigne pas sur sa responsabilité. Il faudrait donc corriger (b'2) en précisant que les mondes possibles pertinents pour évaluer la responsabilité de l'agent sont ceux où il veut faire A dans des conditions normales de délibération ou de décision (ou plus généralement, dans des conditions normales).

Cf. J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 137, et Jeremy Byrd, "Moral Responsibility and Omissions". *The Philosophical Quarterly* 226 (2007), p. 58.

Or il faut faire la même précision à propos du choix que fait S dans le monde actuel, sinon RO* conditionnel pourra attribuer à S la responsabilité de son omission dans certains cas où S est manipulé mentalement⁴⁰. Par exemple, si c'est sous l'influence de l'hypnose que Jean omet de voter pour Luc *dans le monde actuel*, le fait qu'il y ait un monde possible où il vote avec raison pour Luc ne le rend pas responsable de son omission. C'est sans doute la raison pour laquelle McIntyre précise que si l'agent est responsable de son omission, sa décision *actuelle* de ne pas faire A, et sa décision *possible* de faire A sont prises *dans des conditions normales*. Mais cette addition nécessaire pose problème, car il est très difficile de définir ces conditions normales d'un point de vue compatibiliste. Il est possible d'énumérer des exemples de conditions *anormales* (l'agent est drogué, hypnotisé, manipulé, etc.) mais cela n'est pas une définition. Or pour défendre sa position face aux libertariens, le compatibiliste doit prouver qu'il peut définir *complètement* les conditions de la responsabilité morale sans qu'elles excluent la vérité du déterminisme. Tant que ces conditions ne sont pas entièrement définies, il est permis de penser que la responsabilité morale ne se laisse pas définir de manière compatibiliste.

Résumé. Nous avons vu qu'on ne peut définir correctement les conditions de la responsabilité morale dans les omissions de manière conditionnelle. L'obstacle est le même que pour les cas d'action : si un agent a perdu involontairement tout contrôle de ses désirs, et qu'il fait A, il n'est pas responsable de A, même s'il est vrai que *s'il voulait ne pas faire A, il ne ferait pas A*. De même, s'il *ne fait pas A*, il n'est pas responsable de cette omission, même s'il est vrai que *s'il voulait faire A, il le ferait*. Parmi les conditions énoncées dans RO* conditionnel, celles qui attribuent à l'agent un pouvoir des alternatives en le définissant de manière conditionnelle doivent être complétées. Or la seule manière de le faire qui soit en accord avec la thèse compatibiliste est de se référer comme Alison McIntyre à des *mondes possibles* où l'agent agit autrement. Finalement, il est plus simple et plus élégant d'interpréter systématiquement en termes de mondes possibles le mot « pouvoir » dans les conditions de la responsabilité morale. Ceci nous emmène vers la troisième solution compatibiliste pour

⁴⁰ Ou dans un cas où il lui manque un instrument, ou une information indispensable pour agir.

rendre compte des cas d'omissions : celle de Fischer et Ravizza, qui n'utilise plus du tout les formulations conditionnelles.

Au fond, l'analyse conditionnelle des capacités et de la responsabilité morale renvoie de toute façon à des mondes possibles. En effet, une affirmation comme celle-ci :

S est capable de faire A $\leftrightarrow$ si S voulait faire A, il ferait A.

n'aurait guère de sens si S ne pouvait absolument pas *vouloir faire A*. Si cela était logiquement impossible, ou même physiquement impossible (contraire aux lois de la nature), cette définition serait visiblement fausse. Dans le *definiens*, l'antécédent « si S voulait faire A » renvoie donc à une situation possible compte tenu des lois de la nature au moins. Si nous en restons à ce genre de possibilité, l'affirmation peut se reformuler ainsi :

S est capable de faire A $\leftrightarrow$ il y a une situation possible (compte tenu des lois de la nature) où S veut faire A, et fait A en conséquence.

Ainsi, le compatibilisme a tout intérêt à définir les conditions de la responsabilité en termes de mondes possibles. Avant d'examiner cette solution en me penchant particulièrement sur la thèse de Fischer et Ravizza, je voudrais d'abord examiner un argument compatibiliste dont le but n'est pas de réduire les omissions à des actions, mais de minimiser l'asymétrie existant entre ces deux types de cas du point de vue des conditions de la responsabilité.

5.9 Un conflit d'intuitions touchant l'asymétrie entre les actions et les omissions

Dans les sections 5.5 et 5.6, j'ai critiqué la conception réductionniste des omissions, qui nie la spécificité de la responsabilité dans les omissions en réduisant les différents cas d'omissions à des cas d'actions. Certains auteurs ne vont pas aussi loin pour remettre en cause l'asymétrie entre les actions et les omissions. Ils reconnaissent la spécificité des omissions par rapport aux actions, mais soutiennent que les conditions de la responsabilité sont les mêmes dans ces deux sortes de cas. C'est la position de Frankfurt, Fischer, Ravizza, Byrd et Haji⁴¹. Tous admettent que dans « Requins », Jean n'est pas responsable du fait que l'enfant se noie, mais pour eux, cela n'est pas dû au fait que Jean ne peut pas sauver l'enfant⁴². C'est d'abord Frankfurt qui a rejeté cette explication. Il argumente à partir de cet exemple :

[Héros] : Matthieu se promène sur la plage en contemplant la mer. Soudain, il voit un enfant qui se débat dans l'eau pour échapper à la noyade. Il délibère un bref instant, puis se jette à l'eau et sauve l'enfant. En fait, Matthieu ne peut pas agir autrement. Son désir de sauver l'enfant est trop fort pour qu'il reste passif⁴³.

Matthieu est responsable du fait que l'enfant ait échappé à la noyade. Pourtant, il ne peut pas agir autrement. De prime abord, cet exemple pourrait donc aller dans le sens du compatibilisme. Cependant, comme nous l'avons vu dans le premier chapitre, il y a une différence entre la *capacité* et la *possibilité*. Si c'est à cause d'une inhibition ou d'une anomalie psychique que Matthieu ne peut pas agir autrement, on ne peut pas le considérer comme responsable. Mais si, au contraire, l'impossibilité est liée seulement au fait qu'il soit rationnel, il ne s'agit pas nécessairement d'une incapacité, et son cas est analogue à celui du

⁴¹ Harry Frankfurt, "An Alleged Assymetry between Actions and Omissions". *Ethics* 104 (1994), p. 620-623. J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 123-150. Jeremy Byrd, "Moral Responsibility and Omissions". *The Philosophical Quarterly* 226 (2007), p. 56-67. Ishtiyaque Haji, "A Riddle regarding Omissions". *Canadian Journal of Philosophy* 22 (1992), p. 485-502.

⁴² L'argument incompatibiliste fondé sur les cas d'omissions a troublé Fischer et Ravizza, qui ont d'abord admis une « asymétrie » entre les cas d'action positive et les cas d'omission. Dans ceux-ci en effet, la responsabilité morale paraît directement liée à la capacité d'agir autrement. Mais Frankfurt a critiqué cette conclusion, et finalement, Fischer et Ravizza se sont ralliés à sa position. Cf. J. M. Fischer & M. Ravizza, *Responsibility and Control*, p.125, note 3, et p. 127-128.

⁴³ Cet exemple vient de J. M. Fischer et M. Ravizza, "Responsibility and Inevitability," *Ethics* 101 (1991), p. 258-278. Frankfurt le reprend pour montrer que Fischer et Ravizza n'auraient pas dû croire à une asymétrie entre les actions et les omissions.

mathématicien qui ne peut plus faire certaines erreurs à partir du moment où il a acquis la disposition à bien calculer. L'interprétation de cet exemple est donc délicate. Or Frankfurt l'utilise pour interpréter un deuxième cas :

Dans « Paresse », Jean se promène le long d'une plage, et il voit un enfant qui se débat dans l'eau. Jean croit qu'il pourrait sauver l'enfant en faisant un tout petit effort, mais il n'est pas enclin à dépenser la moindre énergie pour sauver quelqu'un d'autre. Il renonce à sauver l'enfant, et continue à marcher le long de la plage.

Jean est-il moralement responsable de ne pas avoir sauvé l'enfant ? A l'insu de Jean, l'enfant était sur le point de se noyer au moment où il l'a aperçu, et l'enfant s'est noyé une seconde après que Jean ait décidé de ne pas se jeter à l'eau⁴⁴.

Jean ne pouvait pas éviter l'accident. Il est blâmable, puisqu'il n'a même pas *essayé* de sauver l'enfant. Mais nous pouvons penser qu'il n'est pas responsable de l'accident. Il est seulement responsable de son choix, et de ne pas avoir *essayé* de sauver l'enfant.

Cette remarque semble d'abord favorable aux compatibilistes. Ils pourraient penser que dans les cas d'omissions, l'agent n'est jamais responsable de l'événement, mais seulement de sa propre attitude. Pourquoi ne penchent-ils pas vers cette solution ? Probablement parce que dans certains cas d'omissions, le droit attribue une responsabilité relative à l'événement lui-même. Si on reprenait l'exemple « Paresse » en imaginant que l'enfant ait la capacité de tenir encore cinq minutes à la surface de l'eau, la conclusion serait différente : Jean serait responsable de l'accident, et nous aurions affaire à un cas classique de non assistance à une personne en danger. Manifestement, le droit ne traite pas l'accusé de la même façon s'il avait ou s'il n'avait pas la capacité d'éviter l'accident : c'est seulement dans le premier cas qu'il est jugé responsable. La responsabilité morale est-elle donc liée à un pouvoir des alternatives ? Pour écarter cette conclusion, Frankfurt imagine encore un exemple :

Dans « Requins », Jean décide de ne pas sauver un enfant qui se noie, et qui se serait noyé même s'il avait essayé de le sauver (il y a en effet des requins aux alentours). Fischer et Ravizza suggèrent qu'il est inapproprié de soutenir qu'en de telles circonstances, Jean est responsable de ne pas avoir sauvé l'enfant.

⁴⁴ J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 125.

Ils ont raison sur ce point. Mais ce qui explique cette discordance n'est pas, comme ils le supposent, le fait qu'il fût impossible à Jean de sauver l'enfant.

Ce fait aurait pu être lié à des circonstances tout à fait différentes de celles qu'ils décrivent. En effet, imaginons que si Jean avait pensé à sauver l'enfant, il aurait été submergé par un désir littéralement irrésistible de faire autre chose ; et imaginons que ceci l'aurait poussé à écarter toute pensée de ce genre. Compte tenu de cette modification, le cas de Jean est exactement parallèle à autre exemple (de Fischer et Ravizza) : celui de Matthieu [Héros]⁴⁵.

Rappelons que dans « Héros », l'agent se précipite à l'eau pour sauver l'enfant, et en raison de sa psychologie, il ne peut pas agir autrement. Dans ce cas, on considère l'agent comme responsable et digne de louange. Frankfurt en déduit que dans le cas où l'agent ne peut pas sauver l'enfant à cause d'une paresse absolument irrésistible, il est également responsable de l'événement. Or contrairement à ce qu'il dit, ces deux exemples ne sont pas strictement parallèles, puisque le premier est un cas d'action bonne, alors que le second est un cas d'action mauvaise. Comme il est *a priori* possible qu'il y ait une asymétrie entre ces deux cas, on ne peut pas s'appuyer sur l'un pour évaluer la responsabilité de l'agent dans l'autre.

Si Jean est proprement *incapable* de réagir de façon appropriée, il faut y voir une sorte de handicap (au sens large) : une cause ou un défaut interne l'empêche de réagir de façon appropriée. Puisque c'est à cause d'un handicap psychologique que Jean ne s'est pas jeté à l'eau, on ne peut sans doute pas le considérer comme responsable de l'accident. Mais on peut expliquer son comportement autrement. C'est peut-être seulement parce qu'il est rationnel (ou vertueux) qu'il se jette à l'eau. Dans ce cas, il n'y a pas de raison de le considérer comme incapable de rester sur la plage, à moins qu'il ne soit limité par une cause interne, psychologique ou nerveuse par exemple⁴⁶. A l'inverse, dans « Paresse », Jean a un comportement irrationnel et on ne peut donc pas l'expliquer en lui attribuant une faculté. Ainsi, dans le premier cas, l'impossibilité pourrait venir d'une faculté ou d'une disposition bonne, alors que dans le deuxième, elle serait due à une incapacité. Cela peut expliquer pourquoi l'agent est responsable dans le premier cas et non dans le deuxième.

⁴⁵ Frankfurt, « An Alleged Asymmetry between Actions and Omissions », *Ethics* 104 (1994), p. 620.

⁴⁶ Dans ce cas, il n'est pas responsable, et l'exemple est en effet parallèle à celui où l'agent est limité par un très fort désir de rester sur la plage (un désir irrésistible).

Frankfurt et Fischer estiment que si, dans « Requins », Jean n'est pas responsable de la noyade, ce n'est pas parce qu'il ne peut empêcher cet événement, mais parce que son environnement n'est pas sensible à ses décisions. Il en découle que dans « Noyade », où il n'y a pas de requins mais un intervenant contrefactuel contrôlant la volonté de l'agent, celui-ci est responsable, car s'il avait choisi de sauver l'enfant, il l'aurait sauvé. Dans ce cas, le sort de l'enfant est sensible aux décisions de l'agent, et c'est pourquoi il en est responsable, selon Frankfurt et Fischer. Cette idée fournirait une manière compatibiliste d'expliquer pourquoi Jean n'est pas responsable dans « Requins ». Cependant, elle aboutit à des conséquences étranges face à des exemples comme celui-ci, imaginé par David Kaplan :

[Requins en cage] Un homme mauvais a mis en cage des requins affamés, et il s'apprête à les relâcher si Jean saute à l'eau. Jean ignore cela. Il voit l'enfant qui se débat dans l'eau, mais décide de ne pas le sauver, et l'enfant se noie.⁴⁷

Pour Fischer et Ravizza, Jean est responsable dans ce cas, alors qu'il ne l'est pas dans « Requins ». En effet, dans « Requins », l'environnement de Jean n'est pas sensible à ses décisions. Quoi qu'il fasse ou décide, les requins sont là. Dans « Requins en cage », au contraire, l'environnement est sensible aux décisions de Jean, car ils ne sont libérés que s'il décide d'aller sauver l'enfant. Fischer et Ravizza soutiennent ainsi qu'il n'y a pas d'asymétrie entre les actions et les omissions ; dans tous les cas, la responsabilité de l'agent serait liée au « contrôle de direction » (*guidance control*), c'est-à-dire à la sensibilité de l'agent aux raisons, et à la sensibilité du monde extérieur aux décisions de l'agent.⁴⁸

⁴⁷ Exemple utilisé par David Kaplan après une conférence de Fischer et Ravizza sur la responsabilité dans les omissions à l'UCLA en février 1995.

⁴⁸ Pour rejeter le « principe asymétrique » et soutenir que la responsabilité est toujours liée au contrôle de direction, Fischer et Ravizza s'appuient aussi sur l'idée que l'omission cause un événement négatif. Si on considère que la responsabilité est relative à un événement positif (par exemple « l'enfant se noie »), il en découle, selon Fischer et Ravizza, que l'omission ne cause pas l'événement dont l'agent est considéré comme responsable. Or le contrôle de direction suppose une relation causale entre l'action ou l'omission de l'agent et l'événement sur lequel il exerce un contrôle. Pour Fischer, l'omission ne se distingue pas de l'action par le fait qu'elle ne cause pas un certain événement positif, mais par le fait qu'elle cause un événement négatif. Cette question est importante pour Fischer, mais elle est sans importance pour nous si nous rejetons son jugement sur l'exemple « Requins en cage ». En outre, j'ai soutenu moi-même que l'omission pouvait causer un événement. Si nous admettons avec Fischer que les omissions causent des événements négatifs, cela ne suffit pas à écarter le principe PPA. En effet, il faut encore comprendre en quel sens elles *causent* ces événements négatifs. Nous avons vu que les exemples de responsabilité dans les omissions manifestent une asymétrie par rapport aux cas

Mais ce jugement est peu intuitif, car la différence entre « Requins » et « Requins en cage » ne paraît pas significative. Je pense qu'ici, Fischer et Ravizza font passer leur théorie avant l'intuition. Il y a au moins trois raisons de le penser. Premièrement, ils reconnaissent eux-mêmes que leur interprétation de « Requins en cage » est loin d'être évidente :

Nous admettons que les cas de ce genre sont difficiles et problématiques. Mais après réflexion, nous maintenons cette distinction et soutenons qu'elle est *justifiée*.⁴⁹

Cette distinction manque effectivement d'évidence et paraît difficile à justifier. C'est pourquoi d'autres compatibilistes ont essayé de défendre le même jugement que Fischer et Ravizza sur la responsabilité de Jean, mais en expliquant autrement la distinction entre « Requins » et « Requins en cage »⁵⁰.

Deuxièmement, si Fischer et Ravizza avaient raison dans leur interprétation du deuxième exemple, il faudrait penser que Jean est responsable aussi dans le premier, car après tout, les requins sont des intervenants contrefactuels : ils n'attaquent Jean que s'il saute à l'eau. Ainsi, dans « Requins », l'environnement de Jean est déjà sensible à ses décisions.

Troisièmement, on peut imaginer des exemples où la différence est encore plus petite qu'entre « Requins » et « Requins en cage », et visiblement sans pertinence pour évaluer la responsabilité de l'agent. Supposons que Jean décide de ne pas sauver l'enfant, mais qu'à son insu, il ne puisse pas le sauver. Il n'y a pas de requin ni d'intervenant contrefactuel, mais un

d'actions, et qu'il est possible de rendre compte de cette asymétrie en soutenant que les omissions causent leurs effets en un sens contrefactuel. Or cette idée implique le principe PPA.

⁴⁹ J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 138.

⁵⁰ Cf. Jeremy Byrd, "Moral Responsibility and Omissions". *The Philosophical Quarterly* 226 (2007), p. 56-67, et Ishtiyaque Haji, "A Riddle regarding Omissions". *Canadian Journal of Philosophy* 22 (1992), p. 485-502. Byrd pense que la différence entre « Requins » et « Requins en cage » est liée à la différence entre des agents non-rationnels et des agents rationnel. Dans le deuxième exemple, l'obstacle est suscité par un individu qui libère les requins. C'est donc seulement dans le deuxième cas que l'obstacle qui empêche Jean de sauver l'enfant est dû à un être doué de raison. Cette distinction proposée par Byrd est aussi peu intuitive que celle de Fischer et Ravizza.

Haji, quant à lui, rejette cette distinction (*ibid.*, p. 494), mais reconnaît son incapacité d'expliquer suffisamment le critère qu'il propose (p. 497). Il pense que l'agent n'est pas responsable dans les cas où l'obstacle est déjà présent avant la décision de l'agent et rend nécessaire l'événement en question. L'agent serait responsable dans le cas contraire, où l'obstacle apparaît seulement si l'agent décide d'empêcher cet événement. La position de Haji est donc voisine de celle que défendent Fischer et Ravizza depuis qu'ils ont abandonné la thèse de l'asymétrie. L'article de Haji (1992), qui critique la thèse de l'asymétrie adoptée dans un premier temps par Fischer et Ravizza, a probablement influencé ceux-ci après leur abandon de cette thèse sous l'influence de Frankfurt.

mur magnétique (invisible) l'empêcherait de nager jusqu'à l'enfant s'il voulait le sauver. Appelons cet exemple « Mur magnétique 1 ». Dans « Mur magnétique 2 », la situation est exactement la même, sauf qu'au lieu d'être présent en permanence, le mur se forme automatiquement si quelqu'un coupe un rayon situé entre Jean et l'enfant. Ainsi, Jean ne peut pas non plus le sauver, car s'il allait vers lui à la nage, il couperait le rayon et serait bloqué par le mur magnétique. La différence entre ces deux cas est insignifiante du point de vue de la responsabilité morale. Dans un cas comme dans l'autre, Jean n'est pas responsable de la noyade. Or, selon le critère de Fischer, il y a un gouffre entre ces deux exemples, car Jean est responsable dans le deuxième et non dans le premier. En effet, dans le deuxième exemple, l'environnement est sensible aux décisions de Jean : c'est seulement s'il va vers l'enfant que le mur magnétique apparaît.

Ces difficultés suggèrent que la distinction proposée par Fischer et Ravizza est une distinction *ad hoc* destinée à éviter la thèse de l'asymétrie et à sauver la thèse compatibiliste. En admettant toutefois qu'elle soit correcte, cela ne suffirait pas pour éviter le problème posé par les omissions. Nous avons vu dans la section 5.8 que pour décrire correctement les conditions de la responsabilité dans les omissions d'un point de vue compatibiliste, il faut les énoncer en se référant à des mondes possibles où l'agent se comporte autrement. Or nous allons voir que ce type de solution ne permet pas de définir les conditions de la responsabilité de façon claire et cohérente.

5.10 Résumé

Lorsqu'un individu est responsable d'un événement par omission, il ne cause pas cet événement dans le même sens que s'il accomplissait une action. De prime abord, sa responsabilité est liée au fait qu'il pouvait faire quelque chose d'efficace contre cet événement. Ainsi, les cas de responsabilité par omission militent en faveur d'un principe voisin de PAP :

(PPA - principe d'action possible) S est responsable de E par omission si E s'est produit et si S avait la possibilité d'agir efficacement contre E.

Après avoir esquissé une théorie libertarienne de la responsabilité dans les omissions (en 5.2), j'ai soutenu (en 5.3) que la différence entre les actions et les omissions n'est pas une simple différence de point de vue. Une omission ne se définit pas comme un comportement passif, car elle peut se réaliser concrètement dans un comportement très actif. Par exemple, on peut omettre de réviser pour un examen en organisant une soirée. De même, un comportement apparemment passif peut être une action à l'égard d'un certain résultat, comme nous l'avons vu dans les exemples « Princesse » 1 et 2 : en restant debout plus longtemps que prévu, la princesse invite ses sujets à la photographier plus longtemps. La notion d'omission est donc une notion relative ; une action à l'égard d'un certain résultat peut être en même temps une omission à l'égard d'un autre résultat. Finalement, nous pouvons retenir les définitions suivantes de l'action et de l'omission :

(DAO - différence entre l'action et l'omission) :

- 1) Un comportement de S qui provoque ou facilite l'événement E est une *action* à l'égard de E.
- 2) Un comportement de S qui ne provoque ni ne facilite l'événement E, mais qui ne s'y oppose pas non plus, est une *omission* à l'égard de E.

En 5.4, j'ai soutenu que l'omission pouvait causer un effet, mais d'une autre manière que l'action. Elle ne produit pas son effet par le biais d'un processus, mais le cause en un sens contrefactuel. Cela permet de comprendre comment la responsabilité d'une personne à l'égard d'une omission se transfère à un événement : la responsabilité d'un agent à l'égard de Y se transmet de X à Y si et seulement si 1°) Y dépend contrefactuellement de X, et 2°) les conditions épistémiques sont satisfaites.

En 5.5, j'ai montré comment cette idée permettait de répondre à une objection fondée sur les cas de surdétermination causale. Dans les sections 5.6 à 5.9, j'ai critiqué les principales positions compatibilistes face aux cas d'omission. Une première solution est de réduire les omissions à des actions (cf. 5.6 et 5.7). Cette solution n'est pas viable pour deux raisons : premièrement, il nous arrive de faire des omissions par inattention. Puisqu'elles ne dérivent pas d'un choix ou d'une action préalable, ces omissions paraissent irréductibles aux actions.

Deuxièmement, il y a des asymétries entre les cas d'omissions et les cas d'actions qui empêchent une telle réduction.

Il faut donc admettre le principe PPA. Mais le pouvoir des alternatives dont PPA suppose l'existence peut être défini de manière compatibiliste : de manière conditionnelle, ou en se référant à certains mondes possibles. En 5.8, j'ai critiqué la première de ces deux solutions, en montrant qu'il faut la compléter ou la remplacer par des formulations en termes de mondes possibles. Supposons un individu S qui omet de faire A, ce qui provoque un événement malheureux. La proposition conditionnelle « si S voulait faire A, il ferait A » n'implique pas la liberté et la responsabilité de S, même si toutes les conditions épistémiques de la responsabilité morale sont satisfaites. En effet, un tel énoncé ne dit pas si S choisit librement de ne pas faire A. Il peut être aliéné jusqu'en ses désirs, par exemple s'il a été drogué, ou manipulé psychologiquement. La vérité du conditionnel est donc compatible avec le fait que S ne soit pas responsable. En outre, le conditionnel ne dit pas si l'action A (ou le choix de faire A) est physiquement possible pour S. Or on ne peut pas attribuer à S le pouvoir de faire A s'il est physiquement impossible qu'il accomplisse cette action. Pour remédier à ce défaut, il faut remplacer la formulation conditionnelle par une formulation en termes de mondes possibles, en précisant de quelles possibilités il s'agit, par exemple si la responsabilité de l'agent dans les omissions suppose la possibilité *à la fois logique et physique* qu'il agisse de manière efficace. Dans le prochain chapitre, j'examinerai en détail les difficultés rencontrées par ce type de solution.

VI La sensibilité aux raisons et les mondes possibles

6.1 Introduction

Nous avons vu dans le chapitre V que la responsabilité par omission n'a pas tout à fait les mêmes conditions que la responsabilité dans les actions. En particulier, les omissions ne peuvent causer leurs résultats qu'en un sens *contrefactuel*. Cela montre la spécificité des omissions et empêche de les réduire à des actions. On s'accorde à dire que dans certains cas d'omissions au moins, être responsable d'un événement par omission suppose le pouvoir de faire une action qui empêche cet événement. Si on complète ce principe en y ajoutant les conditions épistémiques de la responsabilité morale, on se trouve devant trois conditions qui attribuent à l'agent un pouvoir des alternatives. En effet, il est responsable de l'événement E par omission seulement s'il y a au moins une action A telle que :

- (b) il peut faire A.
- (d') il peut savoir que s'il fait A, E ne se produit pas.
- (e') il peut savoir qu'il faut empêcher E.

Ainsi, il y a plusieurs choses que l'agent doit *pouvoir faire* (et qu'il ne fait pas dans la situation actuelle) pour être responsable d'un événement par omission. Or les deux conditions épistémiques que je viens de rappeler s'appliquent également aux actions. En effet, comme nous l'avons vu à la fin du chapitre III, un individu est responsable de E par une action seulement s'il y a une action A telle que :

- il fait A.
- il peut savoir que s'il fait A, E se produira (ou a des chances de se produire).
- il peut connaître la valeur de E (donc savoir s'il faut favoriser ou éviter E).

Le compatibiliste a le choix entre deux manières de définir ces pouvoirs : l'analyse conditionnelle, et l'analyse en termes de mondes possibles. La première ne fonctionne pas, pour les raisons que j'ai exposées plus haut. Dire que « si S veut faire A, il le fait » est insuffisant, car ce conditionnel peut être vrai même dans des cas où l'agent n'est manifestement pas responsable. Par exemple, s'il a été drogué sans le savoir au point de devenir dépendant, il n'est pas responsable du fait qu'il consomme de la drogue, même si on peut dire que s'il voulait s'en priver, il s'en priverait. Clarke et McIntyre supposent qu'on *peut* faire A même si on ne peut pas *vouloir* le faire, mais bien que cette idée soit vraie pour certains genres de capacités, elle ne semble pas s'appliquer aux capacités requises pour la responsabilité morale, car celle-ci est liée au contrôle qu'a l'agent de ses propres actions, et on ne contrôle pas ses propres actions si on ne contrôle pas ses propres choix. Finalement, cette condition a besoin d'un complément : il faut préciser dans une condition supplémentaire que S a le pouvoir de *vouloir faire* A. Or puisqu'il est impossible d'exprimer cela de manière conditionnelle sans tomber dans une régression à l'infini, il faut nécessairement l'exprimer en se référant à des mondes possibles (compris de manière plus ou moins large). Du coup, puisque l'analyse en terme de mondes possibles est indispensable, on peut se demander s'il n'est pas plus simple d'utiliser ce procédé pour toutes les conditions où apparaît le verbe « pouvoir ». On formulerait ainsi les trois conditions (b), (d'), et (e') en termes de mondes possibles.

C'est ce que font Fischer et Ravizza. Selon eux, la condition principale de la responsabilité morale - condition commune aux cas d'actions et aux cas d'omissions - est que le comportement de l'agent résulte d'un « mécanisme sensible aux raisons » (*reasons-responsive*). La notion de sensibilité aux raisons se définit en termes de mondes possibles et signifie à la fois que ce « mécanisme » peut reconnaître des raisons d'agir autrement, et qu'il peut agir selon ces raisons.

Dans ce chapitre, je vais d'abord exposer cette théorie (en 6.2). Ensuite, j'examinerai les difficultés qu'elle entraîne. Une première objection contre cette théorie a été formulée par Stump. Elle se fonde sur des exemples où l'agent fait une mauvaise action, mais où il existe un monde possible où il agit mieux sous l'effet de raisons implantées artificiellement, ou bien grâce à un mécanisme sensible aux raisons qui est modifié (ou implanté) artificiellement. La

discussion de ces exemples (en 6.3) nous emmènera vers une autre objection (en 6.4) : il n'y a pas de critère suffisamment clair, dans une théorie comme celle de Fischer, pour savoir quels sont les mondes possibles pertinents du point de vue de la responsabilité morale. Je le montrerai par une série de contre-exemples. Ensuite (en 6.5), j'exposerai une troisième objection fondée sur les exemples inverses. Ce sont des cas où l'agent est sensible aux raisons (d'après la définition de Fischer et Ravizza), alors qu'il n'est pas responsable dans le monde actuel, à cause d'un facteur excluant la responsabilité morale. Cette objection a été utilisée par Pereboom. Je ferai ensuite (en 6.6) une quatrième objection fondée non plus sur l'idée d'une *cause* excluant la responsabilité de l'agent dans un monde possible ou dans le monde actuel, mais sur un *défaut*, c'est-à-dire un moyen qui manque à l'agent. Enfin, je montrerai (en 6.7) que la sensibilité aux raisons telle que la comprend Fischer n'est pas une disposition (contrairement à ce qu'il déclare). Le concept de sensibilité aux raisons élaboré par Fischer est un concept hybride entre la notion de disposition, et le pouvoir des alternatives compris de manière libertarienne. En effet, l'attribution de la sensibilité aux raisons se fait en fonction du comportement de l'agent dans des mondes possibles proches du monde actuel, alors que l'attribution des dispositions ne dépend pas spécialement des mondes possibles proches du monde actuel. En outre, il y a une frontière nette entre la responsabilité morale et la non-responsabilité, alors qu'il n'y en a pas entre la sensibilité *modérée* aux raisons et la sensibilité *faible* aux raisons. Par conséquent, la notion de sensibilité *modérée* aux raisons ne rend pas compte du pouvoir des alternatives que suppose la responsabilité morale dans les omissions.

6.2 La sensibilité aux raisons

Nous avons vu que dans les cas d'omissions à la Frankfurt, comme « Requins », l'agent n'est pas responsable du résultat, parce qu'il ne peut rien faire d'efficace contre ce résultat. Face à ce genre d'exemples, Fischer et Ravizza ont d'abord pensé que la responsabilité dans les omissions, contrairement à la responsabilité dans les actions, supposait un pouvoir des alternatives au sens libertarien. Cette thèse est souvent appelée « thèse de l'asymétrie ». Elle rejette PAP à propos des actions, mais admet pour les omissions un principe analogue, à savoir PPA (le principe d'action possible). Finalement, Fischer et Ravizza ont abandonné cette

thèse en utilisant la notion de sensibilité aux raisons pour énoncer des conditions de la responsabilité valant à la fois pour les actions et pour les omissions :

S est moralement responsable de son action ou de son omission si et seulement si :

- 1) elle résulte d'un mécanisme appartenant à S, et
- 2) ce mécanisme est sensible aux raisons.

La deuxième condition signifie que l'agent doit pouvoir envisager des raisons d'agir mieux et régler sa conduite en fonction de ces raisons. Selon Fischer et Ravizza, il y a trois types de situations où l'agent n'est pas moralement responsable de ce qu'il fait. Dans chacune de ces situations, la capacité d'agir ou de réagir autrement semble être une condition nécessaire de la responsabilité morale :

La première espèce de défaillance est le fait de ne pas être réceptif aux raisons. Ici, il y a pour ainsi dire des raisons suffisantes d'accomplir l'action, mais l'agent ne reconnaît pas ces raisons. Quand cette espèce de défaillance est due à une incapacité, elle est classiquement associée à la psychose hallucinogène. La seconde espèce de défaillance concerne la réactivité : c'est l'incapacité d'être affecté par ses croyances de façon appropriée. Ici, l'agent reconnaît certaines raisons pour ainsi dire suffisantes pour accomplir une action, mais il ne fait pas ses choix en fonction de cette reconnaissance. Le manque de réactivité affecte certains malades atteints de névrose phobique ou compulsive. Cette espèce de défaillance peut résulter aussi d'une faiblesse de la volonté. Enfin, il y a l'incapacité de traduire pleinement ses choix par des actions ; cette espèce de défaillance peut refléter différentes sortes d'incapacités physiques, ou bien, là aussi, la faiblesse de la volonté¹.

Quelques pages plus loin, Fischer et Ravizza corrigent cette définition sur un point : la faiblesse de la volonté n'est pas un motif suffisant pour excuser l'agent. En effet, la faiblesse de la volonté est un vice dont l'agent peut être responsable, et elle ne l'empêche pas d'agir mieux. Il faut donc admettre une conception moins forte des conditions de la responsabilité morale. Elle est compatible avec l'*akrasia* (faiblesse de la volonté), même si elle est incompatible avec une insensibilité plus forte comme celle d'un psychopathe ayant perdu le

¹ J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 41-42.

contrôle de lui-même. Il faut donc distinguer la sensibilité « forte » et la sensibilité « modérée » :

(SR forte) Pour tout mécanisme K qui aboutit à une action, K est fortement sensible aux raisons si et seulement si : s'il y avait une raison suffisante d'agir autrement, alors l'agent

- 1) reconnaîtrait cette raison,
- 2) choisirait en conséquence d'agir autrement, et
- 3) agirait selon ce choix².

(SR modérée) Pour tout mécanisme K aboutissant à une action, K est faiblement sensible aux raisons si et seulement si il y a au moins un monde possible dans lequel

- 1) les lois de la nature sont les mêmes que dans le monde actuel,
- 2) le mécanisme K est le même, et
- 3) l'agent reconnaît une raison suffisante d'agir autrement, choisit donc d'agir autrement, et agit selon ce choix.

Fischer et Ravizza définissent aussi la sensibilité « faible » aux raisons. La définition est la même que pour la sensibilité modérée, sauf dans la troisième condition, où on ne précise pas que la raison doive être « suffisante ». Une raison superficielle suffit pour la sensibilité faible. Par exemple, si un individu achète de la drogue dans le monde actuel, mais renonce à le faire dans un monde possible où la seule différence avec le monde actuel est qu'il trouve le prix trop élevé, on ne peut pas dire qu'il soit responsable de son action. La sensibilité faible aux raisons est donc insuffisante pour la responsabilité morale.

Notons que la sensibilité forte aux raisons est définie de manière conditionnelle³. Au contraire, la sensibilité modérée (que j'appellerai désormais « sensibilité aux raisons ») ne peut se définir de manière conditionnelle. C'est pourquoi Fischer et Ravizza la définissent en termes de mondes possibles. Précisons d'emblée que ces mondes possibles ne sont pas

² J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 69.

³ Pour plus de précisions sur la sensibilité forte aux raisons, cf. *Responsibility and Control*, p.44, note 16.

accessibles à l'agent, sinon cette condition impliquerait un pouvoir des alternatives au sens libertarien. Dire que l'agent est sensible aux raisons, ce n'est pas dire qu'il a la possibilité d'agir de façon rationnelle au moment et dans les circonstances précises où on le voit agir de façon irrationnelle. Si Fischer et Ravizza concédaient une telle chose, ils seraient contraints d'admettre l'incompatibilité entre le déterminisme et la responsabilité morale, puisqu'elle supposerait l'existence de possibilités alternatives au sens strict. Néanmoins, si la responsabilité ne suppose pas la *possibilité* d'agir de façon rationnelle (au sens le plus fort du mot « possibilité »), elle suppose tout de même la *capacité générale* d'agir de façon rationnelle⁴, ainsi que la capacité particulière d'agir rationnellement dans des situations voisines de la situation actuelle.

6.3 Première objection : un mécanisme ou des raisons implantés artificiellement

Face à cette théorie, la première objection qui vient à l'esprit est que ces conditions n'excluent pas les cas où le mécanisme sensible aux raisons est implanté artificiellement dans le cerveau de l'agent. Imaginons qu'un chirurgien opère un individu à son insu en installant dans son cerveau un petit appareil sensible aux raisons. Avant l'opération, cet homme est impulsif et paranoïaque. Après l'opération, il a une meilleure sensibilité aux raisons, mais son cerveau et sa personnalité ont été profondément modifiés. Cet homme ne paraît pas responsable des choix qui sont provoqués en lui par cet appareil. Mais selon Fischer, l'agent n'est pas responsable si le mécanisme sensible aux raisons ne fait pas partie de lui, et le fait qu'il soit *en lui* ne suffit pas pour qu'il fasse partie de lui. A ce propos, nous pouvons distinguer trois sortes de cas :

Les premiers cas sont ceux où le cerveau de l'agent est modifié par l'installation d'un mécanisme qui altère sa sensibilité aux raisons, en la diminuant ou en l'augmentant. Si cette opération est faite indépendamment de la volonté de l'agent, nous pouvons dire que le nouveau mécanisme sensible aux raisons obtenu par une telle opération n'est pas vraiment

⁴ J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 42-43.

celui de l'agent. Il est donc vain de présenter ce type de contre-exemples. La théorie de Fischer a les moyens d'en rendre compte.

On peut imaginer une deuxième catégorie d'exemples où l'agent subit une opération du même genre, mais parce qu'il l'a voulu. Dans ce cas, il peut être responsable de ses choix, mais seulement de manière dérivée. Puisque par définition, la responsabilité dérivée dérive toujours d'une situation où l'agent est responsable de son choix d'une manière normale et non-dérivée, ce genre d'exemple ne permet pas non plus de réfuter la thèse de Fischer. En effet, Fischer peut répondre que si l'agent est responsable dans ce genre de cas, c'est parce qu'il était sensible aux raisons au moment où il a choisi de subir l'opération.

Enfin, on peut imaginer que le cerveau de l'agent (ou son mécanisme sensible aux raisons) est entièrement remplacé par un autre. Dans ce cas, Fischer estime que l'opération a produit un nouveau « moi » à la place de l'ancien. Par exemple, si je subis cette opération, je cesse d'exister au moment où le nouveau moi est implanté dans mon corps à la place du mien. Une autre personne prend ma place dans ce corps. Dès lors, je ne suis évidemment pas responsable de ce que fait ce nouveau moi. Il s'agit tout simplement d'une autre personne, qui peut être sensible aux raisons comme je l'étais avant l'opération.

Aucun de ces trois genres d'exemples n'est une menace sérieuse pour la théorie de Fischer et Ravizza. Une autre objection plus efficace se fonde sur des exemples où on influence l'agent en lui transmettant des raisons, pour exercer un contrôle sur son comportement sans modifier son cerveau ou ses capacités. Eleonore Stump imagine Raymond, le protagoniste du film *The Manchurian Candidate*, qui a été programmé à l'aide de capteurs de manière à ce qu'il tue son chef. Dans le monde actuel, il capte des raisons qui le poussent à tuer son chef, et il le tue. Or, supposons qu'il y ait un monde possible où les capteurs transmettent à Raymond de bonnes raisons de ne pas tuer son chef. Dans ce monde possible, il reconnaît ces raisons et agit d'après elles. De plus, le mécanisme K est le même dans les deux mondes, comme le demande la théorie de Fischer, car ce qu'on a ajouté, c'est un mécanisme extérieur à l'agent. Raymond est donc sensible aux raisons dans le monde actuel.

Pourtant, il est évident qu'il n'est pas responsable de son action. Le critère de la sensibilité aux raisons n'est donc pas suffisant⁵.

Fischer et Ravizza, eux-mêmes conscients de ce problème, essaient de compléter le critère de la sensibilité aux raisons en disant que le mécanisme doit appartenir à l'agent lui-même⁶. Stump, quant à elle, pense que cette condition supplémentaire ne peut sauver la théorie de Fischer et Ravizza :

Ce n'est pas un remède suffisant pour la théorie de Fischer d'ajouter la condition que le mécanisme soit celui de l'agent lui-même, car vraisemblablement, le mécanisme utilisé par Raymond lorsqu'il tue son chef comme il est programmé à le faire, fait d'une certaine manière partie de lui-même ; c'est son esprit, mais après un lavage de cerveau ou une altération des neurones⁷.

Rappelons en effet que dans cet exemple, l'agent est déjà manipulé dans le monde actuel. On peut ajouter que les lois de la nature sont aussi les mêmes dans ces deux mondes. La seule différence entre la situation actuelle et la situation alternative est le contenu des raisons qui sont communiquées à Raymond, et le comportement de Raymond découlant de cette influence. Pour simplifier l'exemple, nous pouvons imaginer que Raymond soit hypnotisé dans le monde actuel. Supposons qu'un surdoué de l'hypnose l'incite à tuer son chef, et qu'il y ait un monde possible où le même homme hypnotise Raymond avant de lui présenter des raisons de ne pas tuer son chef. Dans le monde actuel, Raymond tue son chef, mais dans cet autre monde possible, il décide de ne pas le tuer, convaincu par les raisons qu'on lui présente. Dans ce cas, le mécanisme est bien le même dans les deux situations et appartient clairement à Raymond. Pourtant, celui-ci n'est pas responsable. Les critères de Fischer et Ravizza ne seraient donc pas suffisants.

Mais selon Fischer, ce genre de contre-exemple suppose une conception trop large du « mécanisme » sensible aux raisons, car dans ce genre de cas, c'est le manipulateur, et non

⁵ Ishtiyaque Haji fait la même objection dans « A Riddle regarding omission » (1992), p. 490.

⁶ Dans *Responsibility and Control*, cette idée fait l'objet d'un long développement à la fin du livre. Nous l'aborderons plus loin avec Pereboom.

⁷ Eleonore Stump, « Intellect, Will, and the Principle of Alternate Possibilities », in *Perspectives on Moral Responsibility*, ed. by J. M. Fischer and M. Ravizza, p. 260.

l'agent, qui est sensible aux raisons⁸. L'agent contrôlé par hypnose ou d'une autre manière n'est pas sensible aux raisons, puisque dans la situation où il se trouve, son choix est déterminé par les raisons qu'on lui transmet *quelles que soient* ces raisons (des raisons spécieuses ou même absurdes lui feraient autant d'effet que de bonnes raisons⁹). Le fait qu'il délibère avant de céder aux raisons communiquées par le manipulateur ne suffit pas pour qu'il soit vraiment sensible aux raisons. En effet, la sensibilité aux raisons n'est pas seulement la capacité de choisir après avoir exercé une activité rationnelle en un sens aussi large, mais la capacité de se laisser convaincre par les *bonnes* raisons plutôt que par les mauvaises. Mais cette réponse suffit-elle ? Je pense que non, car si elle nous éclaire sur la manière dont Fischer conçoit la sensibilité aux raisons, en revanche, elle ne nous montre pas comment la définir en termes de mondes possibles.

De plus, on peut modifier l'exemple en imaginant que les raisons inculquées par le manipulateur ne sont pas irrésistibles, mais très fortes, de telle sorte que Raymond y résiste seulement dans un monde possible où, par exemple, il comprend que s'il tue son chef, cela entraînera la mort de milliers d'innocents. Selon les critères de Fischer et Ravizza, Raymond est sensible aux raisons, puisqu'il y a un monde possible où il reconnaît une raison suffisante pour ne pas tuer son chef. Fischer et Ravizza supposent que dans un tel cas, l'agent n'a pas « pris la responsabilité » du mécanisme d'une manière appropriée. Mais ils admettent finalement qu'il est très difficile de définir ce qu'est une manière appropriée d'en « prendre la responsabilité ». Je ne discuterai pas ce point, car d'autres difficultés plus profondes doivent être examinées, à commencer par celle-ci : en quoi consiste le « mécanisme » qui doit être identique, selon la première condition énoncée plus haut, dans le monde actuel et dans les mondes possibles ? Fischer reconnaît lui-même que sa théorie « ne contient pas de critère explicite de l'individuation du mécanisme », et que ce fait constitue un « problème potentiel »¹⁰. Mais il pense que cette difficulté ne remet pas en question sa théorie des conditions de la responsabilité morale. Je soutiendrai au contraire que cette lacune est le signe

⁸ J. M. Fischer, *My Way*, Oxford University Press, New York, 2006, p. 228-229.

⁹ J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 232.

¹⁰ J. M. Fischer, *My Way*, p. 239.

d'un défaut majeur dans la théorie de Fischer et dans toute théorie compatibiliste fondée sur l'analyse des pouvoirs en termes de mondes possibles.

6.4 Les mondes possibles pertinents. L'impossibilité de trouver un critère

Fischer et Ravizza admettent que la responsabilité suppose toujours des capacités permettant à l'agent de distinguer le Vrai du Faux et le Bien du Mal, puis à traduire cette connaissance en action. Or le contrôle de direction suppose le pouvoir d'agir autrement et d'acquérir certaines connaissances. Quelles sont donc les possibilités alternatives qui appartiennent, selon Fischer et Ravizza, aux conditions de la responsabilité morale ? Pour mieux cerner la difficulté, distinguons trois types de possibilités. On appelle « possible » ce qui est compatible

1. Avec les lois logiques. Ou :
2. Avec les lois logiques et les lois de la nature. Ou bien encore :
3. Avec les lois logiques, les lois de la nature, et le passé actuel.

Le déterminisme exclut l'existence de possibilités alternatives au sens 3, car il voit en tout événement une conséquence nécessaire du passé actuel et des lois de la nature. Ce n'est donc pas le sens 3 qui convient à Fischer et Ravizza. Or ce n'est pas non plus le sens 1, car le critère serait trop souple. Même si l'agent a toutes les maladies psychiques mentionnées par Fischer et Ravizza, il demeure *logiquement* possible qu'il agisse rationnellement, or Fischer et Ravizza considèrent évidemment qu'il n'est pas responsable. Il reste donc la dernière hypothèse : la sensibilité aux raisons se définit en termes de mondes possibles au sens 2. Mais là non plus, la condition n'est pas suffisante. Par exemple, si le comportement irrationnel de l'agent est dû à une lésion cérébrale provoquée par un accident, l'agent n'est pas responsable de ce qu'il fait. Or son handicap n'est pas seulement lié aux lois de la nature ; il est la conséquence d'un événement passé. Où est donc la limite entre les mondes causalement possibles qui sont pertinents et les mondes causalement possibles qui ne le sont pas ?

Fischer et Ravizza soutiennent que la responsabilité de l'agent suppose l'existence d'un « scénario alternatif » dans d'autres mondes possibles où le mécanisme en œuvre est le même. « Mais, bien sûr, la similitude des mécanismes ne requiert pas une similitude dans tous les détails. [...] Rien dans notre conception intuitive d'un mécanisme conduisant à l'action ou dans nos jugements sur des exemples clairs de responsabilité ne nous oblige à dire que la similitude des mécanismes implique une similitude dans les plus petits détails »¹¹. On le concèdera volontiers. Par exemple, si je vole un disque dans un magasin, on peut réfléchir sur les mondes possibles où seule la couleur du papier peint est différente. Si dans ces mondes possibles, je décide finalement d'acheter le disque ou de le remettre en rayon, on peut supposer que dans le monde actuel, très voisin, je suis sensible aux raisons. Il est donc évident que certains détails du monde actuel ne comptent pas. Ainsi, comme le dit Fischer,

il n'y a rien de plausible dans la suggestion que *toutes les conditions du passé* – aussi éloignées et aussi peu pertinentes soient-elles – doivent être comptées comme des parties du « mécanisme aboutissant à l'action ». Si le déterminisme causal est vrai, il doit y avoir un scénario possible (compte tenu des lois de la nature) dans lequel l'agent agit autrement au moment T, et qui diffère en quelque manière du scénario actuel antérieurement à T. L'existence de ces scénarios possibles est tout ce que requiert notre théorie de la responsabilité morale¹².

On voit ici que Fischer et Ravizza ne pensent pas à des mondes possibles au sens 3 (ce qui paraît évident), mais n'adoptent pas non plus le deuxième critère, celui d'un scénario possible au sens 2. S'ils faisaient cela, ils butteraient en effet sur l'objection que j'ai formulée : le critère 2 est trop vague. Pour eux, il faut tout de même que le scénario alternatif soit possible au sens 2, mais ils reconnaissent que cette condition n'est pas suffisante. On se demande alors quelles sont les conditions supplémentaires. Mais ils ne les précisent pas. Et, selon eux, ce flou relatif n'est pas lié à un défaut de leur théorie, mais à la nature même de la responsabilité morale¹³. Cela mériterait des éclaircissements. On peut en effet se demander

¹¹ J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 52.

¹² *Ibid.*, p. 52.

¹³ « La réflexion sur des cas comme les animaux intelligents, les psychopathes et les enfants, met en lumière la difficulté de produire une analyse claire et nette de la responsabilité morale. En effet, bien que SRR et WRR [la sensibilité forte et la sensibilité faible aux raisons] soient des formulations assez précises, elles sont toutes les deux inadéquates. Nous avons apporté quelques corrections à WRR. Bien que ces corrections la rendent plus proche de la vérité, elles la rendent aussi beaucoup moins précise. C'est regrettable à certains égards, mais nous

quel est le sens de cette remarque. Signifie-t-elle que la distinction entre les mondes possibles pertinents et les autres n'est pas nette, les différents mondes possibles étant *plus ou moins* pertinents, ou bien que la différence est nette mais apparaît seulement au cas par cas, ce qui empêcherait d'énoncer une règle générale pour les distinguer ?

S'ils veulent dire que la différence apparaît seulement au cas par cas, c'est un aveu d'échec, car ils ont le devoir de formuler une définition aussi claire et opérationnelle que les définitions incompatibilistes¹⁴. Cependant, pour Fischer et Ravizza, il n'y a peut-être pas de limite précise entre les possibilités pertinentes et les autres. Il en serait alors comme pour les autres dispositions (telle la solubilité du sucre). Par exemple, si l'agent agit mieux que dans le monde actuel lorsqu'il est sous l'influence de sa mère ou en sa présence, il a une certaine sensibilité aux raisons. On lui attribuera donc une responsabilité, limitée mais non négligeable. S'il n'agit mieux que sous l'influence d'un individu qui le sermonne tout en le menaçant ou en le soumettant à un chantage, il a encore une sensibilité aux raisons, bien qu'elle se manifeste plus rarement et dans des situations plus éloignées du monde actuel. Il n'aurait donc qu'une responsabilité très faible, vu la faiblesse de sa disposition à la rationalité.

croyons que c'est un résultat inévitable de la complexité et de la confusion des phénomènes de la responsabilité eux-mêmes. Notre pratique des attributions de responsabilité est complexe et elle-même indéterminée, c'est pourquoi il n'est pas surprenant que les concepts et les distinctions employés dans la théorie de la responsabilité morale reflètent jusqu'à un certain point cette indétermination. » (*Responsibility and Control*, p. 80-81)

¹⁴ Fischer et Ravizza reconnaissent volontiers que la notion la plus commune de la responsabilité suppose des possibilités alternatives, mais ils veulent montrer qu'il est possible de la redéfinir en abandonnant ce présupposé. Ils doivent donc nous proposer une définition suffisamment claire pour que nous sachions si elle ne requiert plus l'existence de possibilités alternatives. Or c'est justement ce qu'ils ne font pas. En évoquant des mondes possibles où l'agent se comporte autrement, ils reviennent à une idée qui rappelle les possibilités alternatives. Nous devons donc savoir si cette idée est cohérente, ce qu'on ne peut faire sans la préciser entièrement.

Fischer et Ravizza proposent une définition de la responsabilité prétendument compatible avec le déterminisme, mais au moment où on leur demande de préciser cette définition sur un point où elle paraît incompatible avec le déterminisme, ils nous laissent face aux mêmes difficultés qu'avant, car en précisant que le scénario alternatif doit être compatible avec les lois de la nature et avec *certains* aspects du passé actuel (mais pas tous), ils nous laissent face aux problèmes qu'ils avaient l'obligation de résoudre : il faudra se demander quels aspects du passé actuel sont vraiment pertinents. S'agit-il de facteurs externes à l'agent, ou bien de causes internes, par exemple ? Sûrement pas de causes externes qui détermineraient le comportement de l'agent. S'agit-il plutôt de causes internes ? Non plus, car si l'action est déterminée par les propriétés génétiques de l'agent, elle n'est pas accomplie librement, etc. Le problème reste donc entier, et en l'absence de réponse vraiment précise, la théorie de Fischer et Ravizza n'augmente pas notre compréhension de la responsabilité morale, car en elle-même, l'idée que la responsabilité morale suppose une aptitude à la rationalité, est banale. Ce qu'ils doivent montrer, c'est que cette aptitude ne suppose pas de contingence au sens le plus fort de ce mot.

Pour examiner la pensée de Fischer et Ravizza sur ce point, il faut tester leur théorie en imaginant concrètement des mondes possibles plus ou moins proches du monde actuel. Je discuterai d'abord des cas où l'agent subit l'influence d'une *cause* externe ou interne *dans certains mondes possibles voisins du monde actuel*. Je soutiendrai que l'existence de tels mondes possibles ne nous renseigne pas sur la responsabilité de l'agent. Ensuite, je ferai l'objection inverse, c'est-à-dire fondée sur des cas où l'agent subit l'influence de *causes* externes ou internes *dans le monde actuel*. L'existence de mondes possibles où il agit mieux ne prouve pas qu'il soit responsable et sensible aux raisons s'il est influencé par une de ces causes dans le monde actuel. Enfin, je ferai des objections analogues à propos des cas où l'agent est limité par un *obstacle* ou un *défaut*.

Pour simplifier, je partirai toujours de la même situation. Supposons qu'il arrive ceci dans le monde actuel :

(Disque 1) Pierre flâne dans les magasins un samedi après-midi et il tombe sur le dernier album de son chanteur préféré. La tentation est forte. Il voudrait l'emporter tout de suite, mais n'a pas d'argent pour l'acheter. Il décide donc de le voler. Quand Pierre atteint la sortie du magasin, le disque n'est pas détecté par le système antivol, mais Pierre est arrêté par un membre du personnel qui l'a vu ôter l'emballage du disque alors qu'il était dans les rayons.

D'après Fischer et Ravizza, pour savoir si Pierre est moralement responsable, il faut savoir s'il est sensible aux raisons, ce qu'on ne peut décider autrement qu'en examinant des situations possibles assez proches de la situation présente. Commençons par une situation où Pierre est influencé par une cause interne :

(Disque 2) Pierre a subi une opération du cerveau. Un appareil implanté dans son cerveau l'empêche de céder à la tentation lorsqu'il désire voler quelque chose. Au moment où il se prépare à voler le disque, l'appareil déclenche un signal nerveux qui le fait changer d'avis.

Fischer et Ravizza eux-mêmes disent assez clairement pourquoi l'existence d'un tel monde possible ne fournit pas un indice pertinent. Comme nous l'avons vu plus haut, le *moi*

de l'agent ne doit pas être altéré. « Disque 2 » n'est pas assez proche du monde actuel, car les capacités de l'agent sont nettement différentes dans les deux cas. Mais qu'en est-il si nous introduisons la plus petite modification possible à l'intérieur de l'agent ?

(Disque 3) Dans « Disque 3 », la série d'événements est la même que dans « Disque 1 » jusqu'au moment du choix, à un détail près : il y a une minuscule partie de son cerveau qui est différente. La configuration des neurones à cet endroit fait que le comportement de Pierre n'est pas exactement le même : il hésite, réfléchit, puis repose le disque au lieu de le voler.

Cette modification est sans doute la plus petite qu'on puisse imaginer (ce peut être une différence au niveau d'un seul neurone). Or il n'y a qu'une différence de degré, non de nature, entre cette minuscule différence et une autre plus importante affectant les capacités générales d'une partie du cerveau. Changeons un seul neurone de Pierre, ou un seul de ses états neuronaux. La modification n'est pas grande. Pourtant, la réaction de Pierre dans ce monde possible ne nous renseigne nullement sur sa responsabilité dans le monde actuel. En effet, aussi petite que soit la modification du point de vue morphologique ou physiologique, son importance se mesure à ses effets dans le comportement de l'agent. Si elle lui permet d'agir autrement, c'est qu'elle a modifié sensiblement le mécanisme par lequel il réagit aux raisons. Il faut donc apporter cette précision sur le second critère de Fischer, celui de l'identité du mécanisme : le mécanisme sensible aux raisons est le même seulement si la différence de comportement par rapport au monde actuel ne vient pas d'une modification *à l'intérieur de l'agent*, par exemple dans son cerveau. Il faut donc que les mécanismes internes à l'agent qui interviennent dans le monde actuel existent aussi dans le monde possible ; autrement, on n'apprend rien sur la sensibilité aux raisons du mécanisme actuel. Mais encore faut-il savoir jusqu'où s'étend le « mécanisme ». Il ne peut pas inclure tous les éléments du monde actuel qui jouent un rôle dans le processus causal aboutissant à l'action, car d'un point de vue déterministe, une telle identité implique que le résultat est le même dans les deux cas, un ensemble donné de causes produisant toujours le même effet.

La seule issue est de dire que les modifications permises portent sur des aspects *extérieurs* à l'agent, et qui influencent son comportement en le poussant à agir mieux. Nous pouvons donc retenir ce principe : pour savoir si l'agent est sensible aux raisons dans un cas où il n'agit pas de façon rationnelle, il faut imaginer des mondes possibles où il agit de façon rationnelle mais où cette différence n'est pas due à une différence interne à l'agent. Les mondes possibles pertinents sont des mondes possibles où la différence de comportement par rapport au monde actuel est due à une modification dans l'environnement de l'agent. Mais le problème qui se pose alors est que dans toutes ces situations, l'agent est influencé par une cause extérieure, et il perd sa responsabilité dans la mesure où il est influencé. C'est ce que nous allons voir dans les prochains exemples.

(Disque 4) La situation est la même que dans « Disque 1 », sauf qu'un individu influence les pensées de Pierre grâce à un appareil permettant d'interrompre et de provoquer à distance la transmission de messages nerveux. Au moment où Pierre se prépare à voler le disque, l'individu intervient en déclenchant un désir soudain de laisser le disque en rayon et de se racheter en accomplissant une action vertueuse, par exemple, passer tout son dimanche après-midi en compagnie de sa grand-mère.

Dans cet exemple, le cerveau et les mécanismes internes à l'agent sont les mêmes que dans le monde actuel. Pourtant, le changement introduit dans le passé actuel n'est pas pertinent, car il fait que Pierre est manipulé par quelqu'un. Son action ne prouve pas une véritable sensibilité aux raisons. Mais on peut imaginer des mondes possibles où Pierre subit une influence qui n'est pas irrésistible :

(Disque 5) Au moment où Pierre est sur le point de déballer le disque et d'en enlever la pastille adhésive conçue pour être détectée par le système antivol, un inconnu lui fait observer ce qu'il y a d'égoïste et d'immoral dans le vol qu'il est en train de commettre. En même temps, il le menace de le dénoncer. Pierre revient à la raison et laisse le disque à sa place.

Dans cette situation, Pierre agit mieux que dans le monde actuel, mais sous la menace. La réaction de Pierre dans ce monde possible est donc trop ambiguë pour nous renseigner sur son degré de sensibilité aux raisons.

Imaginons alors ceci :

(Disque 6) Comme dans « Disque 5 », Pierre se trouve devant un inconnu qui essaie de le ramener à la raison, mais cette fois-ci, l'homme ne le menace pas. Pierre a honte. Il le remercie et laisse le disque en rayon.

Le fait que Pierre agisse comme il faut dans ce monde possible est-il une preuve de sa sensibilité aux raisons et de sa responsabilité ? Nous pouvons bien lui attribuer un certain degré de sensibilité aux raisons, mais un enfant serait tout aussi responsable que lui. En effet, si un enfant fait une mauvaise action, on ne le considère pas comme *pleinement* responsable de ce qu'il fait sous prétexte qu'il y a des mondes possibles où il renonce à cette action sous l'influence de sa mère. Inversement, lorsqu'un enfant accomplit, sous l'influence d'un adulte, une action qu'il n'accomplirait pas autrement, il n'est pas entièrement responsable de ce qu'il fait¹⁵. Ainsi, le changement introduit dans « Disque 6 » par rapport au monde actuel nous empêche de le considérer comme *pleinement* responsable.

Nous avons dit un peu plus haut que pour Fischer et Ravizza, la sensibilité aux raisons n'était pas un pouvoir des opposés donnant accès à différents mondes possibles dans une même situation donnée¹⁶. En cela, elle est comparable aux dispositions des plantes et des animaux : pour savoir si une plante est capable de réagir de telle ou telle façon, on se réfère à d'autres situations ou mondes possibles. De même, pour savoir si un individu est sensible aux raisons, il faudrait se référer à d'autres mondes possibles. Soit un guépard domestique nommé Brutus. Il suit la jeep de son maître lors d'une pointe à 90 km/h. On se demande s'il est capable de courir plus vite. Il faut alors savoir ce qu'il ferait si la jeep roulait à 110 km/h, ou

¹⁵ Il arrive qu'un individu choisisse de façon pleinement autonome en suivant l'avis de quelqu'un, mais lorsqu'un individu fait son choix d'après l'avis d'une autre personne et qu'il ne ferait pas le même choix en son absence, nous estimons qu'il est influencé par elle (dans la situation actuelle) et qu'il n'est pas pleinement autonome.

¹⁶ Nous prenons ici le mot « possibilité » au sens le plus fort. Sur cette question, Fischer et Ravizza sont explicites : « Ainsi, nous avons associé la responsabilité morale à une propriété modale ou dispositionnelle. Il est important de voir que même si d'autres mondes possibles sont pertinents pour déterminer si une certaine disposition (comme la sensibilité faible aux raisons) opère ou non dans le monde actuel, ces mondes ne sont pas pertinents pour traiter la question de savoir si telle ou telle séquence alternative est *véritablement* (*genuinely*) accessible à l'agent. » (*ibid.*, p. 53).

s'il a déjà atteint cette vitesse en poursuivant une antilope¹⁷. Il est évident qu'on ne pourra pas trancher en imaginant une situation où le mécanisme est différent, par exemple, en imaginant l'animal plus jeune et plus grand. Les seules modifications permises théoriquement portent donc sur l'environnement extérieur, comme dans notre exemple, et la même observation s'applique aux dispositions des plantes et des choses inertes. Pour savoir si un corps est soluble ou inflammable, il ne faut pas l'imaginer avec des propriétés chimiques différentes, mais procéder à une sorte d'expérience de pensée en le plaçant mentalement dans d'autres situations ou interviennent des facteurs extérieurs. Il est donc clair que les modifications pertinentes introduiront des causes extérieures influençant le comportement de l'agent. Et puisque l'action de telles causes enlève la responsabilité morale, la théorie de Fischer et Ravizza est dans une impasse, à moins de renoncer à une intuition très forte. Nous verrons que dans un monde possible où l'agent agit mieux, la cause qui le fait agir mieux constitue toujours une différence trop grande par rapport au monde actuel. Quelle que soit cette cause, elle enlève toute pertinence à ce monde possible dans l'évaluation de la responsabilité de l'agent.

Je vais maintenant explorer deux options possibles pour la thèse compatibiliste. Soit elle considère qu'il y a une limite nette entre les mondes possibles pertinents et les autres, soit elle estime qu'il n'y en a pas, et que tout monde possible où l'agent agit mieux que dans le monde actuel a une certaine pertinence pour évaluer sa responsabilité. Je soutiendrai que dans les deux cas, le critère de la sensibilité aux raisons se révèle insuffisant.

¹⁷ Pour savoir si le guépard peut courir à 110 km/h, il faut savoir si on peut concevoir un monde possible où il court plus vite. Rappelons que ce monde possible ne doit pas seulement être logiquement possible ; il doit aussi, selon Fischer, être proche du monde actuel. Donc pour répondre à la question, on peut s'appuyer sur des observations déjà faites dans le passé : par exemple, il a déjà atteint 110 km/h deux semaines plus tôt dans des circonstances semblables.

Y a-t-il une limite nette entre les mondes possibles pertinents et les autres ? Première option

Soit le compatibiliste soutient qu'il y a une limite nette entre les mondes possibles pertinents et les autres, soit il estime qu'il n'y en a pas. La première option (qui n'est pas celle de Fischer et Ravizza) semble conduire à une régression à l'infini. En effet, dans tous les mondes possibles que nous avons imaginés, nous sommes tombés exactement sur les mêmes difficultés que dans le monde actuel, car les seuls mondes possibles pertinents semblent être ceux où l'agent agit mieux que dans le monde actuel, et où il est en outre *responsable de ce qu'il fait*. Un tel critère nous entraîne dans une régression à l'infini. En effet, au demeurant, il n'y a aucune raison de penser qu'on puisse s'arrêter à un monde possible où les mêmes problèmes théoriques ne se poseront plus. Par exemple, pour savoir si l'agent est responsable dans « Disque 1 » (le monde actuel), on se réfère à « Disque 5 ». Mais pour que ce monde possible nous renseigne sur la sensibilité aux raisons de Pierre, il faut savoir s'il y agit sous l'effet de la menace, ou parce qu'il est sensible aux bonnes raisons formulées par l'individu qui le menace. Le problème se pose autant qu'au départ. Si on se référait plutôt à « Disque 6 », le problème se poserait également. En effet, savons-nous vraiment s'il y a un monde possible tel que « Disque 6 » où Pierre agit mieux que dans le monde actuel *du fait de sa sensibilité aux raisons* et non pas seulement parce qu'il est très influençable ? Chaque monde possible pose des problèmes analogues : à chaque fois, on peut se demander si l'agent accomplit son action par sensibilité aux raisons ou sous l'effet d'un autre facteur. Cette difficulté ne peut être surmontée sans un critère qui délimite exactement le domaine des possibilités pertinentes.

Dans l'esprit de Fischer et Ravizza, la sensibilité aux raisons s'analyse seulement à partir de *ce que fait* l'agent dans certains mondes possibles, et non à partir des *pouvoirs ou capacités* qu'il a dans ces mondes possibles. Or si nous faisons abstraction de ces pouvoirs ou capacités, il nous faudra tout de même un critère pour délimiter les mondes possibles pertinents. Les exemples que j'ai passés en revue montrent que la *cause* du comportement de l'agent dans ces mondes possibles n'est pas indifférente. Par exemple, si dans un monde possible, on introduit dans l'agent lui-même une cause qui le fait agir autrement (que dans le monde actuel), on altère le « mécanisme » qui fait partie de l'agent et qui aboutit à son action

ou à sa décision. Pour compléter le critère de la sensibilité aux raisons, et finalement pour qu'il fonctionne comme critère, il faudrait donc distinguer deux catégories parmi les causes qui déterminent les comportements humains, en montrant quelle catégorie de causes n'exclut pas la responsabilité morale.

Deuxième option : la sensibilité aux raisons étant une disposition, il n'y a pas de limite nette entre les mondes possibles pertinents et les autres

La comparaison avec les dispositions volontaires ou involontaires suggère une réponse à l'objection : aucun monde possible n'est parfaitement pertinent pour estimer la responsabilité morale de l'agent, mais tous les mondes possibles où les propriétés de l'agent ne sont pas modifiées et où il agit de façon rationnelle prouvent dans une certaine mesure sa sensibilité aux raisons, aussi faible soit-elle. En somme, la sensibilité aux raisons serait proportionnée à la quantité de mondes possibles où il agit rationnellement. Par exemple « Disque 6 », où Pierre renonce à voler le disque après avoir écouté les arguments de quelqu'un, et même « Disque 5 » où il agit de la même façon, mais sous la menace, prouvent sa sensibilité aux raisons ; seulement cette disposition est faible si ce sont les seuls mondes possibles où il renonce à voler le disque.

Fischer et Ravizza estiment que lorsqu'on se réfère à des mondes possibles pour attribuer à l'agent la sensibilité aux raisons, on ne peut pas dire précisément quels aspects du passé actuel ne doivent pas changer. Mais ils déclarent que cette difficulté est liée à la nature même de la responsabilité¹⁸. S'il y a une différence nette entre ceux qui sont pertinents et ceux qui ne le sont pas (première option), le critère de la sensibilité aux raisons est incomplet et pose le problème que je viens de mettre en évidence. Mais pour Fischer et Ravizza, la sensibilité aux raisons est une propriété dispositionnelle, et comme les autres dispositions volontaires ou involontaires, elle admet des degrés. Cela signifierait que plus il y a de mondes possibles où l'agent agit rationnellement, plus il possède cette disposition ; même si Pierre n'agit rationnellement que dans des cas très particuliers, il possède une sensibilité aux raisons,

¹⁸ Cf. J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 80-81.

aussi faible soit-elle. Ainsi, dans le cas où Pierre est menacé (« Disque 5 »), le fait qu'un tel monde possible existe suffirait pour lui attribuer une *sensibilité aux raisons*. Le problème est que cela ne paraît pas suffisant pour lui attribuer la *responsabilité* de ce qu'il fait. Dans un tel cas, la différence entre le monde possible et le monde actuel est déjà trop grande, bien qu'il soit difficile d'en imaginer une moins grande. Ainsi, toute différence avec le monde actuel permettant de concevoir Pierre agissant bien nous paraît trop grande pour lui attribuer la responsabilité morale.

Pour toute disposition, il y a des mondes possibles clairement pertinents et d'autres qui le sont moins, sans différence nette entre eux. La sensibilité aux raisons serait donc comparable aux autres dispositions, par exemple à celles des plantes et des animaux. Si un guépard ne dépasse les 100 km/h que rarement, lorsqu'il est en pleine forme, nous pouvons sans doute dire qu'il a la capacité de courir à plus de 100 km/h, mais sa disposition à atteindre une telle vitesse est faible, beaucoup plus faible que s'il l'atteignait une fois par semaine. Une disposition est faible lorsque celui qui la possède n'agit selon elle que rarement, c'est-à-dire dans un petit nombre de mondes possibles. Et une disposition est forte dans le cas inverse. Si Pierre n'agit rationnellement qu'en présence d'un adulte ou d'une personne veillant sur lui, sa sensibilité aux raisons est faible, et s'il faut toujours le menacer pour qu'il écoute sa raison, elle est encore plus faible.

Finalement, puisque la sensibilité aux raisons est une disposition de l'agent et une disposition *de sa volonté*, ne doit-elle pas être comprise comme une vertu ? Rappelons que pour Fischer et Ravizza, une sensibilité aux raisons *modérée* suffit pour être pleinement responsable de ce qu'on fait. La sensibilité aux raisons est donc une disposition volontaire, mais contrairement aux vertus, elle appartient à tout individu normal, et en dessous d'un certain seuil, elle est trop faible pour que l'agent soit responsable de ses actions. Mais où placer cette limite ? Là encore, le critère de la sensibilité aux raisons se révèle insuffisant.

Puisqu'il y a une limite nette entre la responsabilité et la non-responsabilité, il faut qu'une limite nette sépare les mondes possibles pertinents et les mondes possibles non-pertinents pour l'attribution de la sensibilité modérée aux raisons. Dans tous les exemples que nous avons vus précédemment, la cause qui fait que l'agent choisit mieux dans le monde

possible en question est d'une telle nature que ce monde possible ne paraît pas pertinent pour évaluer la responsabilité morale de l'agent. Il faudrait donc compléter le critère de la sensibilité aux raisons en montrant quel genre de cause peut amener l'agent à agir mieux dans certains mondes possibles sans que ceux-ci nous paraissent hors de propos. Un compatibiliste pourra toujours établir une liste de facteurs excluant la pertinence d'un monde possible. Par exemple, il posera que si S fait A (une action mauvaise) dans le monde actuel, il n'est pas responsable si, dans les seuls mondes possibles où il ne fait pas A, cette différence par rapport au monde actuel s'explique par le fait que :

- S est drogué.
- S est manipulé psychologiquement (hypnotisé, menacé, etc.).
- Le cerveau de S a été modifié artificiellement.
- S a reçu une meilleure éducation morale que dans le monde actuel.

Mais si une telle liste peut sauver la thèse incompatibiliste face aux différents exemples qu'on lui oppose, elle ne donne pas une véritable compréhension de la nature de la responsabilité morale. Pour rendre compte de nos jugements sur la responsabilité morale des agents dans un cadre compatibiliste, il faut un critère positif. Une liste de contre-exemples ne satisfait pas cette exigence minimale. Dans la prochaine section, je soutiendrai qu'un problème analogue se pose en ce qui concerne le monde actuel. Le critère de la sensibilité aux raisons est insuffisant (tel qu'il est présenté par Fischer) parce qu'il peut être satisfait dans des situations actuelles où l'agent est drogué, manipulé, etc. Pour compléter les conditions énoncées par Fischer, il faudrait donc ajouter que l'agent ne doit pas être drogué, manipulé (etc.) *dans le monde actuel*.

6.5 L'objection inverse : un agent sensible aux raisons, mais qui ne se contrôle pas dans le monde actuel

Dans la section précédente, j'ai critiqué le recours compatibiliste aux mondes possibles en présentant des exemples de situations *non actuelles* où l'agent agit mieux que dans le monde actuel, mais sous l'influence d'une cause qui semble disqualifier le monde possible en question (l'agent y est hypnotisé, manipulé psychologiquement, etc.). Or il est également possible d'utiliser la stratégie inverse : imaginer des situations *actuelles* où l'agent n'est *pas responsable* parce qu'il choisit sous l'effet d'une certaine cause qui l'empêche de se contrôler. Le fait qu'une telle cause agisse sur lui ne l'empêche pas nécessairement d'être sensible aux raisons, car cette faculté implique seulement qu'il existe un monde possible voisin du monde actuel où il agit mieux parce qu'il a vu une bonne raison de le faire. Derk Pereboom a appliqué cette méthode pour critiquer la thèse compatibiliste. Il présente une série de cas correspondant à tous les critères compatibilistes :

Dans chacun des quatre cas que j'expose, le professeur Plum décide de tuer Ms White en vue d'un avantage personnel, et y parvient. Cet acte homicide est conforme aux principales conditions compatibilistes que l'on désigne comme suffisantes pour la responsabilité morale d'un agent lorsqu'on y ajoute certaines précisions tout à fait indiscutables. Cet acte satisfait les différentes conditions proposées par Hume et ses successeurs : il est causé par des désirs découlant de son caractère « durable et constant », puisque pour lui, les raisons égoïstes ont généralement un grand poids – beaucoup trop grand si on les juge d'un point de vue moral, mais le désir d'après lequel il agit n'est pas irrésistible pour lui, et en ce sens il n'est pas contraint d'agir (Hume 1739/1978 : 319-412). Il s'accorde avec la condition proposée par Frankfurt : le désir de Plum de tuer White est conforme à ses désirs de second ordre (...) au sens où il veut la tuer et veut sa volonté de le faire, et il veut cet acte homicide parce qu'il veut sa volonté de le faire (Frankfurt 1971). L'action rejoint la condition de sensibilité aux raisons proposée par Fischer et Ravizza (...). Elle satisfait aussi la condition proposée par Jay Wallace : Plum garde sa capacité générale de maîtriser, diriger, et régler son comportement à l'aide de raisons morales. Par exemple, lorsque des raisons égoïstes défavorables à un comportement moral sont relativement faibles, il règlera généralement son comportement selon des raisons morales.¹⁹

¹⁹ D. Pereboom, R. Kane, J. M. Fischer, et M. Vargas, *Four Views on Free Will*, Blackwell, Malden (Massachusetts), 2007, p. 94.

Voici maintenant les exemples de Pereboom. Dans « Plum 1 », le professeur Plum est manipulé par des scientifiques au moyen d'un système radio. Son action a toutes les caractéristiques énoncées par Pereboom dans le passage que je viens de citer. Or nous ne le jugeons pas responsable, car il a tué sous l'effet d'une manipulation. Dans *Responsibility and Control*, Fischer et Ravizza veulent rendre compte de ce genre d'exemples. Ils complètent ainsi leur définition de la responsabilité : le mécanisme sensible aux raisons doit faire partie de l'agent lui-même ; l'agent doit en « prendre la responsabilité ». Ce processus demande du temps et consiste en trois choses :

Premièrement, un individu doit se voir lui-même comme la source de son comportement au sens que nous avons indiqué. Autrement dit, l'individu doit se voir lui-même comme un agent ; il doit voir que ses choix et actions ont un effet dans le monde. [...]

Deuxièmement, l'individu doit admettre qu'il est à juste titre une cible d'attitudes réactives, vu la manière dont il agit dans certains contextes.²⁰

[...] La troisième condition de la prise de responsabilité est que la représentation de soi-même que nous avons définie dans les deux premières conditions, soit fondée d'une manière appropriée sur l'évidence.²¹

La notion de responsabilité morale est une notion « historique », au sens où elle fait intervenir des considérations sur les origines plus ou moins éloignées de l'événement ou de l'acte dont on veut savoir si l'agent en est responsable. Selon Fischer et Ravizza, cette dimension « historique » est liée au fait que l'agent met un certain temps à endosser la responsabilité des ressorts de ses actions. Pour savoir s'il est responsable, il faut remonter en arrière dans son histoire personnelle. Face à des exemples tels que « Plum 1 », Fischer et Ravizza peuvent donc répondre que l'agent fait l'objet d'une manipulation localisable dans le temps, sans que nous ayons besoin pour cela de remonter très loin en arrière. En examinant l'histoire de l'agent, on arrive assez vite à un moment où des individus lui implantent quelque chose dans le cerveau. Cette précision donne à Fischer et Ravizza le moyen de répondre à l'objection. Pour eux, c'est parce que l'agent n'a pas pris la responsabilité des ressorts de son action qu'il n'en est pas responsable.

²⁰ J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 211.

²¹ *Ibid.*, p. 213.

Dans « Plum 2 », Pereboom modifie l'exemple afin de radicaliser l'objection. La manipulation est la même, mais commence cette fois dès la naissance de l'agent. Or la conclusion est la même. « Le fait que le programme prenne place deux secondes ou trente ans avant l'action paraît sans pertinence du point de vue de la responsabilité morale ».²²

Dans « Plum 3 », la manipulation change de nature : l'agent est déterminé par des pratiques d'embrigadement utilisées dans son milieu. Cet entraînement commencé très tôt fait qu'il est souvent (mais non exclusivement) un égoïste rationnel. Vu l'âge qu'il avait au début de cette éducation, il n'avait pas la capacité d'y résister. Sa tendance égoïste est donc très forte, bien qu'elle ne s'impose pas dans tous les cas. Plum tue White, il veut tuer White, il veut sa volonté de le faire, etc. Il est sensible aux raisons car il y a des mondes possibles où il trouve des raisons de pas le faire et où il renonce au meurtre à cause de ces raisons. Pourtant, on ne le considère pas comme responsable moralement. Si nous remplaçons enfin l'embrigadement par un déterminisme purement physique ou biologique, la conclusion n'est pas différente, selon Pereboom, l'action étant toujours déterminée par des événements qui échappent au contrôle de l'agent.

La seule issue est de dire que dans « Plum 3 », Plum n'est pas responsable parce que l'action est causée par d'autres agents que Plum lui-même. Il faut donc comparer ce cas à un dernier exemple où cette cause intelligente est remplacée par une cause aveugle et purement physique :

Imaginons un cas supplémentaire, qui est exactement le même que les cas 1 et 2, sauf que les états de Plum sont induits par une machine qui est générée spontanément, sans dessein intelligent. Le compatibiliste peut admettre que cette sorte de machine à induction détruit aussi bien la responsabilité, et inventer alors une condition stipulant que les agents ne sont pas responsables pour des actions manipulées par des agents ou des machines. Cependant, cette manœuvre semble *ad hoc* – elle paraît motivée uniquement par le désir d'aboutir à un résultat compatibiliste. A ce point de la discussion, le compatibiliste pourrait essayer de tracer une ligne quelque part entre la manipulation par un agent et l'induction venant

²² D. Pereboom, R. Kane, J. M. Fischer, et M. Vargas, *Four Views on Free Will*, p. 96.

d'une machine, mais je ne vois pas comment cette manœuvre pourrait être développée d'une manière plausible.²³

En somme, l'objection de Pereboom est que si Plum n'est pas responsable dans « Plum 1 » ou dans « Plum 2 », on ne voit pas pourquoi il serait responsable dans « Plum 4 ». Le déterminisme causal serait donc incompatible avec la responsabilité morale, même lorsque l'agent possède la sensibilité aux raisons. Fischer répond de façon inattendue que Plum est déjà responsable dans les premiers exemples²⁴. Il n'y aurait donc pas à chercher une ligne de démarcation, parmi les quatre exemples, entre ceux où l'agent est responsable et ceux où il ne l'est pas. Dans un premier temps, Fischer ne se prononce pas sur « Plum 1 » et « concède, pour les besoins de la discussion » que Plum est manipulé dès sa naissance dans Plum 1 (comme dans Plum 2). Dans ce cas,

Plum a pris la responsabilité du mécanisme de manipulation ; après tout, c'est le mécanisme par lequel il agit toujours, et lorsqu'un individu se développe pour devenir un agent moralement responsable, il prend la responsabilité du mécanisme actuellement en œuvre²⁵, même s'il en ignore le détail.²⁶

Selon Fischer, Plum, dans le deuxième exemple, a « pris la responsabilité » du processus de manipulation, et il faut donc le considérer comme responsable de ce qu'il fait, car Pereboom a précisé que le désir provoqué en Plum par les manipulateurs n'est pas irrésistible. Il s'ensuit que Plum agit par lui-même, selon un mécanisme sensible aux raisons. Mais en disant cela, Fischer justifie son jugement à partir de sa théorie. Nous nous trouvons ainsi face à un conflit d'intuitions.²⁷

Que peut répondre Fischer si dans « Plum 1 », l'agent n'a pas « pris la responsabilité » du processus de manipulation ? Dans sa réponse à Pereboom, Fischer a modifié cet exemple en l'assimilant à « Plum 2 ». Ce faisant, il a contourné une difficulté importante, car dans

²³ *Ibid.*, p. 96.

²⁴ J. M. Fischer, *My Way*, p. 232.

²⁵ “*actual-sequence mechanism*”.

²⁶ *Ibid.*, p. 232.

²⁷ Fischer argument aussi en disant que si Plum ne paraît pas responsable à première vue dans les deux premiers exemples, c'est seulement parce qu'il n'est pas blâmable (le contrôle qu'il a sur ses actions étant trop limité). Cf. *Ibid.* p. 233.

« Plum 1 » tel qu'il est présenté par Pereboom, l'agent ne semble pas responsable. Or s'il n'est pas responsable dans « Plum 1 », il paraît encore moins responsable dans « Plum 2 », où la manipulation commence dès la naissance de l'agent, modifiant encore plus profondément sa personnalité. Et s'il n'est pas responsable dans « Plum 2 », pourquoi le serait-il dans « Plum 3 » et « Plum 4 » ? Faut-il supposer que tout agent humain normal prend la responsabilité du déterminisme causal ? Il est difficile de répondre à cette question sans une définition plus précise de la « prise de responsabilité ». Rappelons que pour prendre la responsabilité d'un processus, l'agent doit se voir lui-même comme un agent rationnel et une cible potentielle d'attitudes réactives, mais il faut aussi que la « représentation de soi-même » ainsi définie « soit fondée d'une manière appropriée sur l'évidence »²⁸. En l'absence de cette précision, nous pourrions objecter que l'agent déterminé causalement peut se tromper lorsqu'il se représente lui-même comme un agent rationnel et libre. La précision donnée par Fischer écarte cette objection, mais on peut se demander si elle ne nous entraîne pas dans un cercle ou dans une régression à l'infini. En effet, pour savoir que l'agent ne se trompe pas en se voyant lui-même comme un agent rationnel et responsable de ce qu'il fait, il faut déjà connaître les critères de la responsabilité morale.

Une objection voisine de celle de Pereboom a été formulée par Ishtiyaque Haji. Son exemple est le suivant :

[Drogue] Annie est dépendante de l'héroïne. Elle sait que si elle se retient de prendre de l'héroïne à un certain moment, elle va mourir. Compte tenu de son addiction, Annie prend de la drogue à ce moment particulier. Or elle agit de même dans tous les situations possibles sauf une, où elle choisit courageusement de se retenir pour prolonger la vie du jeune Frank, qui a également besoin de drogue.²⁹

Annie n'est pas responsable du fait qu'elle prend de la drogue, car son addiction est si forte qu'elle ne se contrôle pas elle-même. Supposons qu'elle ait été droguée contre son gré jusqu'au moment où elle est devenue dépendante. Dans ce cas, elle n'est pas responsable de sa dépendance, et dans l'exemple de Haji, elle n'est pas responsable du fait qu'elle prenne de la drogue, même d'une manière dérivée. Pourtant, elle est sensible aux raisons, puisqu'il y a un

²⁸ J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 213.

²⁹ I. Haji, « A Riddle regarding Omissions » (1992), p. 490.

monde possible où elle s'en abstient pour une excellente raison. Pour tenir compte de ce genre de cas, Fischer devrait ajouter une condition supplémentaire : l'agent est responsable de ce qu'il fait seulement s'il choisit de le faire *dans des conditions normales de délibération*. Nous retrouvons ainsi la condition (b) d'Alison McIntyre, que j'ai discutée dans le chapitre V. Le problème est le même ici : comment définir ces « conditions normales » ? Les compatibilistes peuvent expliquer *négativement* de quoi il s'agit en donnant des exemples de conditions anormales : l'agent n'est pas responsable de son choix s'il le fait sous l'influence de la folie, d'une drogue, de l'hypnose, d'une menace ou d'une manipulation quelconque. Mais cela ne constitue nullement une définition. Tant que cette définition manque, rien ne permet de dire que le déterminisme est compatible avec la responsabilité morale, car pour en être sûr, il faut nécessairement donner une définition claire et complète des conditions de la responsabilité qui soit compatible avec le déterminisme causal.

6.6 Défauts ou obstacles dans le monde actuel ou dans les mondes possibles

Dans les sections précédentes, j'ai défendu l'idée que la sensibilité aux raisons ne pouvait pas se définir de manière compatibiliste par la référence à certains mondes possibles. Or nous avons vu à la fin du chapitre V et au début du chapitre VI que le compatibilisme exige un tel recours. Je pense donc que la thèse compatibiliste est dans une impasse lorsqu'il s'agit de rendre compte des pouvoirs de l'agent liés à la responsabilité. J'ai soutenu cette idée à partir d'exemples où l'agent subit une influence aliénante dans le monde actuel ou dans les mondes possibles où il agit mieux. Je vais maintenant exposer un argument analogue mais fondé sur des cas où l'agent est limité par un défaut. Considérons d'abord le cas suivant :

[Asphyxie] Vers 2h du matin, Jean est réveillé par des cris. Son immeuble est en train de brûler. Il se lève et cherche à sauver son voisin de pallier, dont la porte est fermée à clef de l'intérieur. Ce vieillard est trop sourd pour entendre les cris ; il continue à dormir. Or les pompiers ne sont pas encore là, et Jean n'a pas les clefs de son voisin. Il voudrait enfoncer la porte, mais après deux ou trois essais, il voit qu'elle est trop solide. Il cherche enfin un pied de biche ou une barre de fer, mais aucun objet ne lui paraît suffisant pour tenir lieu de pied de biche. La fumée envahit de plus en plus les couloirs.

Jean est obligé de descendre au plus vite en renonçant à sauver le vieillard, et celui-ci meurt asphyxié deux minutes après.

Jean n'est pas responsable de la mort de son voisin. Pourtant, il y a des mondes possibles où celui-ci n'a pas fermé sa porte à clef. Il y en a aussi où il l'a fermée à clef, mais a donné un double à Jean, pour les cas d'extrême urgence. Enfin, il y a des mondes possibles où Jean trouve un pied de biche ou un instrument suffisant pour forcer la porte. Dans ces mondes possibles, Jean voit des raisons de sauver son voisin (comme dans le monde actuel), et il le fait, car la porte n'est pas pour lui un obstacle. Ainsi, Jean est sensible aux raisons selon les critères de Fischer. Pourtant, il n'est pas responsable de la mort de son voisin. Nous avons vu comment Fischer interprète ce genre de cas. Comme dans « Requins », l'agent n'est pas responsable du résultat car un *obstacle* l'empêche d'éviter ce résultat (et pour cette raison, l'environnement de Jean n'est pas sensible à ses décisions). Cela revient à dire que Jean n'est pas responsable parce qu'il lui manque les moyens nécessaires au sauvetage de son voisin. En effet, comme nous l'avons vu dans le premier chapitre, la présence d'un obstacle équivaut généralement à l'absence d'un moyen nécessaire pour contourner ou éliminer cet obstacle. Cette absence de moyen est ce qu'on peut appeler un défaut *externe*.

Le défaut ne va pas toujours de pair avec un obstacle physique. En effet, un simple défaut *d'information* peut rendre l'agent incapable d'agir mieux. A moins qu'il ne soit responsable de sa propre ignorance, il n'est pas responsable de ce qu'il fait *par ignorance*. Voici un cas de ce genre :

[Syndicat] Jean et Pierre font partie du même syndicat. Un nouveau responsable syndical va être élu lors de la prochaine réunion. Jean hésite à voter pour Luc. Finalement, il vote pour Marc, qui est élu de justesse. Par la suite, Marc s'avère un très mauvais représentant syndical. Il est plus soucieux de sa carrière personnelle que du bien commun des employés. Or il y a un monde possible où Jean vote pour Luc au lieu de Marc. Dans ce monde possible, un collègue de Jean lui donne une information très confidentielle sur la personnalité de Marc. Jean n'aurait pu deviner de telles choses. Cette information prouve que Marc ne mérite pas la confiance des employés. Jean ne vote donc pas pour lui, mais pour Luc.

Jean vote pour Luc dans le monde actuel, ce qui est un mauvais choix. Mais est-il responsable de son vote ? Non, car il ne peut pas soupçonner Marc d'être un individu aussi peu fiable. Pourtant, il est sensible aux raisons, et d'une manière générale, un compatibiliste dira qu'il peut voter autrement, car il y a des mondes possibles où il le fait. De plus, dans ces mondes possibles, il le fait pour une excellente raison, car il connaît les intentions profondes de Marc. Il est donc clair que l'existence de ces mondes possibles ne nous dit rien sur la responsabilité de Jean dans le monde actuel.

Un compatibiliste peut cependant répondre en complétant le critère de la sensibilité aux raisons par les conditions épistémiques suivantes :

S est responsable de l'action A seulement si

(a) S *sait* qu'il fait A

(b) S *sait* s'il peut faire A ou s'il doit éviter de faire A.

Ces conditions peuvent paraître trop strictes car nous sommes parfois responsables d'une action ou d'une omission sans savoir ce que nous devons faire. C'est le cas lorsque nous sommes responsables de notre propre ignorance. Par exemple, si je manque une réunion professionnelle importante pour avoir omis d'ouvrir mon courrier ou de consulter un panneau d'affichage, je suis responsable de mon absence à cette réunion, car je savais qu'en négligeant de faire cela, je risquais de manquer une information importante. La solution la plus naturelle pour rendre compte de ce genre de cas d'un point de vue compatibiliste est de les traiter comme des cas de responsabilité dérivée. Dans l'exemple « Syndicat », Jean serait responsable de son vote s'il avait choisi librement (ou négligé volontairement) de se renseigner comme il le devait sur les candidats au poste de représentant. S'il avait su, même de manière vague, qu'il risquait de se tromper gravement en omettant de s'informer d'avantage, il serait responsable de son vote de manière dérivée. Dans le cas présenté plus haut, Jean se trouve au contraire dans une situation d'ignorance « invincible », selon l'expression traditionnelle des théologiens : il ne peut pas savoir, ni deviner vaguement qu'il doit agir autrement.

En ce qui concerne les conditions épistémiques de la responsabilité, le compatibilisme ne rencontre pas de difficulté particulière, mais le critère de la sensibilité aux raisons, même complété par ces conditions, ne suffit pas, puisqu'il peut être satisfait sans que l'agent soit responsable. Nous avons vu dans la section 6.4 que l'agent peut être sensible aux raisons selon le critère de Fischer tout en étant manipulé dans le monde actuel. Et nous venons de voir qu'il peut aussi être sensible aux raisons sans avoir les moyens requis pour agir mieux. Tout cela montre que le critère de la sensibilité aux raisons est insuffisant. Je reviens donc à la question cruciale déjà posée plus haut : comment compléter ce critère ? Le compatibiliste pourra toujours le compléter en énonçant une liste de facteurs qui ne doivent pas intervenir dans le monde actuel. Par exemple, il pourra spécifier que S n'est pas responsable si :

- Il est drogué
- Il est manipulé psychologiquement (hypnotisé, menacé, etc.)
- Tout son entourage lui a inculqué des valeurs morales radicalement fausses
- etc.

Pour les cas d'omission, il ajoutera que S n'est pas responsable de son omission consistant à ne pas faire A si un obstacle l'empêchait de faire A.

Mais une telle liste de contre-exemples (qu'on peut allonger *ad libitum*) n'est pas une définition et ne nous donne pas une vraie compréhension de la notion de responsabilité. Recourir à une telle liste est une solution *ad hoc*. En outre, elle rend inutile le critère de la sensibilité aux raisons. En effet, si cette liste est assez complète pour couvrir tous les cas de non-responsabilité (ou de non-sensibilité aux raisons), elle n'est plus un simple complément du critère de la sensibilité aux raisons, mais remplace celui-ci.

Résumé. Nous avons vu qu'une formulation compatibiliste des conditions de la responsabilité morale qui analyse le pouvoir des alternatives en termes de mondes possibles pose un problème de clarté. En effet, il est très difficile de délimiter l'ensemble des mondes

possibles pertinents pour savoir si l'agent est sensible aux raisons et moralement responsable. Selon Fischer, ils sont *logiquement* et *physiquement* possibles, et dans ces mondes possibles, le « mécanisme sensible aux raisons » doit être le même que dans le monde actuel. Or si le déterminisme est vrai, le mécanisme ne peut être exactement le même sans aboutir au même résultat. C'est pourquoi le critère de l'identité du mécanisme ne peut porter sur tous les détails du mécanisme sensible aux raisons. Les mondes possibles où l'agent agit autrement que dans le monde actuel en diffèrent nécessairement par l'intervention d'un facteur qui n'existe pas dans le monde actuel. En 6.4, j'ai utilisé divers exemples pour montrer que dans les mondes possibles où une différence dans les causes du comportement de l'agent le fait agir autrement que dans le monde actuel, cette différence, quelle qu'elle soit, paraît trop grande pour que ces mondes possibles nous renseignent sur la responsabilité de l'agent. Pour compléter le critère de la sensibilité aux raisons, il semble donc nécessaire de préciser quels facteurs imaginés dans les mondes possibles ne s'opposent pas à la pertinence de ces mondes possibles.

Or en 6.5, nous avons vu qu'un problème analogue se pose concernant le monde actuel. Un individu peut être sensible aux raisons selon le critère de Fischer et Ravizza tout en étant manipulé (et non responsable) dans le monde actuel. Le fait qu'il agisse mieux dans des situations voisines du monde actuel ne permet donc pas d'affirmer sa responsabilité morale. Il faut aussi que *dans le monde actuel*, il agisse sans être manipulé, drogué, etc. Là encore, donc, il faut tenir compte des causes ou des facteurs qui le font agir. Ainsi, l'analyse de la responsabilité en termes de mondes possibles se révèle incomplète des deux côtés : du côté des mondes possibles, et du côté du monde actuel. De part et d'autre, il est nécessaire de la compléter en tenant compte des facteurs dont découle le choix ou l'action.

En 6.6, nous avons vu que les facteurs à prendre en compte ne sont pas seulement les causes qui provoquent le choix ou l'action, mais aussi les obstacles et les défauts, car un obstacle physique, un défaut d'information, ou le manque d'un moyen nécessaire peuvent s'opposer à la responsabilité morale.

Finalement, ces considérations font apparaître une relative indépendance entre la sensibilité aux raisons et la situation réelle de l'agent dans le monde actuel : l'agent est sensible aux raisons s'il agit mieux dans des situations *proches du monde actuel*, mais ce

critère est vague et semble autoriser une telle séparation, comme on le voit dans la manière dont Fischer interprète les exemples de Pereboom. Cette indépendance est un défaut de la théorie compatibiliste, car il est peu intuitif de penser que le pouvoir des alternatives lié à la responsabilité morale soit sans rapport avec la situation de l'agent dans le monde actuel. Lorsqu'une cause ou un facteur exclut la responsabilité morale de l'agent, c'est justement cette cause ou ce facteur qui lui enlève son pouvoir des alternatives. Les pouvoirs que suppose la responsabilité morale ne sont pas séparables des conditions causales dans lesquelles nous agissons. Pour éviter cette séparation, il faut en venir à une compréhension libertarienne du pouvoir des alternatives, selon laquelle l'agent agit mieux dans au moins un monde possible, mais dans un monde possible *identique* par ailleurs au monde actuel. Dans le chapitre III, nous avons vu comment ces deux aspects sont liés entre eux. Un homme hypnotisé ou drogué (ou même un homme qui ignore ce qu'il doit faire) peut perdre ainsi le pouvoir d'agir autrement ; or il l'aurait encore si ce pouvoir se définissait en termes de mondes possibles non accessibles dans la situation actuelle. Pour qu'il y ait un lien entre les conditions causales où se trouve l'agent, et son pouvoir des alternatives, il faut que les mondes possibles pertinents du point de vue de la responsabilité morale soient compris comme *réellement accessibles à l'agent*, ce qui exclut le déterminisme causal.

Du point de vue compatibiliste, le pouvoir des alternatives ne peut exister que sous la forme d'une disposition. Dans la théorie de Fischer et Ravizza, par exemple, la sensibilité aux raisons est une propriété dispositionnelle. Dans la prochaine section, je vais prolonger les critiques précédentes en examinant quels liens peuvent exister entre la responsabilité morale et les dispositions de l'agent.

6.7 Cinquième objection : si la sensibilité aux raisons est une disposition, on doit pouvoir l'attribuer à l'agent sans se référer à des mondes possibles proches du monde actuel

Pour rendre compte du pouvoir des alternatives que nous possédons lorsque nous sommes responsables d'un événement par omission, le compatibilisme a besoin d'analyser ce pouvoir en termes de mondes possibles. Ce genre de définition s'applique normalement aux

dispositions, c'est pourquoi Fischer et Ravizza comprennent la sensibilité aux raisons comme une propriété *dispositionnelle*. Or nous allons voir qu'elle n'en est pas une, car l'attribution d'une disposition ne se fait pas seulement en fonction de mondes possibles proches du monde actuel. Je soutiendrai que ce concept élaboré par Fischer et Ravizza est un concept hybride qui ne peut exprimer les conditions de la responsabilité morale.

La présence d'une disposition est connue à partir d'autres situations où elle passe à l'acte. Nous savons qu'un morceau de sucre est soluble parce que nous avons vu d'autres morceaux de sucre se dissoudre, dans d'autres situations. En disant qu'un morceau de sucre est soluble dans l'eau, nous signifions que si on le plongeait dans l'eau, il se dissoudrait. De la même façon, en attribuant à l'agent la sensibilité aux raisons, nous voulons seulement dire, selon Fischer, que s'il était placé dans telle ou telle situation, il agirait de façon rationnelle.

La sensibilité aux raisons est une *bonne* disposition. Puisqu'elle est bonne, elle ne justifie pas à elle seule un jugement négatif sur l'agent. On ne peut pas reprocher à quelqu'un d'avoir tendance à bien agir dans certaines situations. Si un cheval manque un obstacle alors qu'il le passe bien depuis un mois, il est décevant, et le cavalier a raison d'être furieux contre lui, mais il ne dira pas que ce cheval est moins bon que celui qui rate *toujours* les obstacles. De même, si un individu agit en général de façon vertueuse puis accomplit tout d'un coup une mauvaise action, on ne lui reprochera pas d'agir mieux dans d'autres circonstances. La seule chose qu'on lui reprochera, c'est l'action qu'il aura accomplie, et s'il a de meilleures dispositions qu'un autre agent, il est meilleur. Un compatibiliste peut résoudre cette difficulté en soutenant que la culpabilité naît d'un contraste entre une action *mauvaise* et une *bonne* disposition. Examinons cette hypothèse.

Il est certain que nous tenons parfois compte des dispositions de l'agent lorsque nous évaluons sa responsabilité. Par exemple, il n'est pas pleinement responsable de ce qu'il fait s'il agit sous l'influence d'une mauvaise disposition acquise involontairement. S'il est puissamment attiré par la drogue après avoir été drogué de force ou sans le savoir, on ne peut pas lui reprocher de céder à son penchant. Mais à part ce genre de cas, l'agent est responsable de ses propres dispositions mauvaises. Comme l'a écrit Aristote,

S'il dépend de nous d'accomplir les actions bonnes et les actions honteuses, et pareillement encore de ne pas les accomplir [...], il en résulte qu'il est également en notre pouvoir d'être intrinsèquement vertueux ou vicieux.³⁰

[...] et si, sans avoir l'ignorance pour excuse, on accomplit des actions qui auront pour conséquence de nous rendre injustes, c'est volontairement qu'on sera injuste.³¹

Puisque la sensibilité aux raisons est une bonne disposition de la volonté, et qu'une disposition de ce genre est une vertu, nous pouvons dire que l'agent est *vertueux* s'il possède une forte sensibilité aux raisons, *vicieux* s'il en possède une faible, et *normal* s'il en possède une modérée. Selon Fischer, l'agent n'est pas responsable de ce qu'il fait s'il ne possède au moins une sensibilité *modérée* aux raisons. Un tel critère est problématique, car il semble ignorer le fait qu'on est généralement responsable de ses vices. Cependant, le compatibiliste pourra toujours répondre que lorsqu'un individu est responsable de ses vices (et de la dégradation de sa sensibilité aux raisons), c'est parce qu'il les a acquis dans des situations où il avait encore une sensibilité aux raisons suffisante (donc *modérée*). Cette disposition ferait donc partie des conditions de la responsabilité dans les situations normales, c'est-à-dire en dehors des cas de responsabilité dérivée.

Il y a cependant une grande différence entre le critère de la sensibilité aux raisons et celui que nous appliquons lorsque nous attribuons les autres dispositions. En effet, pour savoir si un individu possède une vertu, nous n'avons pas besoin d'envisager des mondes possibles proches du monde actuel. Or selon Fischer, la sensibilité aux raisons ne s'évalue pas seulement en termes de mondes possibles, mais en termes de mondes possibles *proches du monde actuel*. Un homme est sincère s'il dit la vérité et évite le mensonge dans la plupart des situations, notamment celles où il est difficile d'agir ainsi. Pour savoir s'il est sincère, nous n'avons pas besoin d'imaginer des mondes possibles particulièrement proches du monde actuel. Il est même plus éclairant d'envisager des situations très différentes. Supposons que dans le monde actuel, la tentation du mensonge soit très faible pour l'agent. Si nous voulons mesurer sa sensibilité aux raisons, nous avons intérêt à imaginer des situations où la tentation

³⁰ Aristote, *Ethique à Nicomaque* III, 7, p. 140-141 (traduction J. Tricot).

³¹ *Ibid.*, p. 143.

est beaucoup plus forte, et d'autres où la tentation est beaucoup plus faible. Il faut aussi imaginer des tentations variées. En somme, *l'attribution d'une disposition se fait à partir de l'éventail le plus large de mondes possibles.*

Pourquoi le critère de la sensibilité aux raisons ne fonctionnerait-il pas de la même façon ? Pourquoi envisager des mondes possibles proches du monde actuel, en laissant de côté les autres mondes possibles ? Nous procédons ainsi parce que nous voulons savoir ce que l'agent a le pouvoir de faire *dans le monde actuel*. Autrement dit, ce n'est pas la disposition elle-même que nous cherchons à connaître, mais le pouvoir de l'agent de faire une *autre action* dans des conditions *identiques* à celles du monde actuel.

Au fond, ce que Fischer appelle la sensibilité modérée aux raisons ne ressemble pas vraiment à une disposition. En effet, pour savoir si un agent a une disposition D à un degré normal, nous considérons des mondes possibles plus ou moins variés, mais sans accorder une importance particulière à ceux qui sont proches du monde actuel. En outre, le degré *normal* de cette disposition ne varie pas d'une situation *actuelle* à une autre. Or dans la théorie de Fischer, la sensibilité *modérée* aux raisons est plus ou moins forte selon les cas. Rappelons en effet que pour Fischer, la sensibilité *modérée* aux raisons est le degré minimal de sensibilité aux raisons que doit posséder l'agent pour être responsable. Or, selon les situations, l'agent a besoin d'une sensibilité aux raisons plus ou moins forte pour être responsable. Si on me torture cruellement pour que je révèle une information importante et confidentielle, j'ai besoin d'une forte sensibilité aux raisons pour ne pas céder. Au contraire, si on veut me faire torturer quelqu'un en échange d'une petite somme d'argent, je n'ai besoin que d'une faible sensibilité aux raisons pour refuser cette proposition. Si nous comprenons la sensibilité aux raisons comme une disposition, nous devons penser que la responsabilité morale ne suppose pas toujours cette disposition au même degré. Ainsi, l'agent n'aurait pas toujours besoin d'une sensibilité *modérée* aux raisons. Si dans la théorie de Fischer, la responsabilité morale suppose toujours une sensibilité *modérée* aux raisons, c'est parce que, tout en déclarant qu'il s'agit d'une disposition, Fischer la définit autrement qu'une disposition : il postule en effet que l'attribution de la sensibilité modérée aux raisons se fait à partir de mondes possibles proches du monde actuel.

La sensibilité aux raisons, dans la théorie de Fischer, est donc un concept hybride : un intermédiaire entre les dispositions proprement dites, et le pouvoir des alternatives compris au sens libertarien. La leçon que nous pouvons tirer de cette analyse est que le compatibilisme ne parvient pas à fonder l'attribution de la *responsabilité* morale sur l'attribution de *dispositions* morales. Pour attribuer la responsabilité morale, nous avons besoin d'envisager des situations proches du monde actuel, voire identiques à celui-ci jusqu'à l'instant du choix. Le plus naturel est d'interpréter cela de manière libertarienne, en attribuant à l'agent un pouvoir des alternatives au sens le plus fort.

En généralisant cette critique à l'analyse compatibiliste du pouvoir des alternatives, nous pouvons présenter ainsi les trois notions que je viens d'évoquer : la sensibilité aux raisons (et plus généralement le pouvoir des alternatives compris de manière compatibiliste) est un concept intermédiaire entre la notion de disposition et le pouvoir des alternatives compris de manière libertarienne :

- **Dispositions** : S a une disposition à faire A $\leftrightarrow$ S fait A dans la plupart des mondes possibles où il a l'opportunité de faire A.
- **PA compatibiliste** : S peut faire A $\leftrightarrow$ S fait A dans des situations possibles proches de la situation actuelle.
- **PA libertarien** : S a le pouvoir de faire A à t $\leftrightarrow$ S fait A à t dans des situations identiques à la situation actuelle avant t.

Comme nous l'avons vu plus haut³², tous les mondes possibles sont plus ou moins pertinents pour l'attribution d'une disposition, et il n'y a pas de limite nette entre une disposition forte et la même disposition possédée à un degré plus faible. Au contraire, il y a une différence nette entre la responsabilité et la non-responsabilité. La frontière entre les mondes possibles pertinents et les mondes possibles non-pertinents pour attribuer la responsabilité morale doit donc être nette, et ne peut correspondre, comme le suppose Fischer, à la différence entre la sensibilité *modérée* et la sensibilité *faible* aux raisons. Entre avoir et ne

³² Cf. *supra*, p. 291.

pas avoir *du tout* une certaine disposition, il y a une différence nette, mais non entre la sensibilité modérée et la sensibilité faible aux raisons.

En élaborant le concept de sensibilité aux raisons, Fischer et Ravizza cherchent une voie médiane entre des intuitions relatives aux dispositions et d'autres relatives au pouvoir des alternatives compris de manière libertarienne. On peut reprocher à quelqu'un d'avoir une *mauvaise disposition*, ou de ne pas avoir telle ou telle bonne disposition. On peut également lui reprocher de ne pas avoir agi mieux s'il pouvait agir mieux *dans le monde actuel*. Mais il semble incohérent de mélanger ces deux intuitions.

Finalement, du point de vue compatibiliste, il est peut-être plus cohérent de fonder la responsabilité d'un mal sur une *mauvaise* disposition. Mais une telle solution est profondément contre-intuitive. Si un cheval a tendance à manquer les obstacles, il est tout simplement mauvais pour le saut d'obstacle ; mais dira-t-on de même que certains individus sont *mauvais pour la moralité*, et doivent être punis à ce titre ? Non, car ce raisonnement est contraire à nos intuitions. D'abord, on ne punit personne pour une mauvaise *disposition*, mais pour une mauvaise *action*. Ensuite, la présence d'une mauvaise disposition dans l'agent est même souvent une circonstance atténuante. Par exemple, un criminel qui ne peut se refréner est jugé moins sévèrement que s'il en était capable. Lorsqu'un crime s'explique par une disposition si forte que le criminel avait peu de chance d'y résister, on ne s'y réfère pas pour l'accuser, mais pour l'excuser et atténuer sa peine.

Pour illustrer cette opposition, j'évoquerai une scène fameuse du film de Fritz Lang *M le Maudit*. A la fin du film, le tueur psychopathe doit se défendre face à un tribunal insolite composé d'une foule de bandits. En assassinant des enfants, l'accusé a obligé la police à sortir de sa passivité coutumière. Elle s'est engagée dans une enquête systématique dans les lieux suspects de la ville, ce qui empêche les hommes des mafias locales d'exercer leur métier tranquillement. Dans la scène du procès, le tueur se défend en expliquant que lorsqu'il voit un enfant, il ne peut résister au désir de le suivre et de le tuer. Beaucoup y voient une circonstance atténuante, mais d'autres y voient une circonstance aggravante :

- Qui me croira ? Qui sait ce qui se passe en moi ? Ces cris, ces hurlements...cette force qui me pousse. Je résiste, elle me pousse ! Je résiste, elle me pousse encore ! Puis cette voix stridente...Je n'en peux plus de l'entendre ! Au secours ! Je n'en peux plus.

- L'accusé vient de nous dire qu'il est obligé d'assassiner. C'est-à-dire qu'il y est poussé... Il a prononcé sa sentence de mort. Un homme qui avoue être obligé de détruire des vies humaines, il faut l'abattre comme une bête fauve ! Il faut l'exterminer ! Il faut l'éliminer !

Le gangster qui prononce cette accusation a une conception inhabituelle de la responsabilité morale : il voit dans l'incapacité d'éviter le mal un défaut ou une sorte de vice justifiant le recours à la peine de mort. En fait, il ne s'appuie pas du tout sur la notion de responsabilité morale. C'est du moins ce que suggère la comparaison avec une « bête fauve ». On trouve quelque chose d'approchant dans un passage de Leibniz :

il faut quelquefois ôter des membres pourris, mauvais quand même leur mal ne dépendait point d'eux, comme on ôte du troupeau une brebis galeuse, comme on enferme un furieux.³³

L'incapacité de faire le bien peut-elle être un motif d'accusation ? Cette idée est étrange, mais peut-être plus cohérente que les définitions de la responsabilité morale proposées par les compatibilistes. Cela dit, l'inconvénient attaché à cette thèse est plus grand que son avantage : elle est simple et cohérente en elle-même, mais profondément contraire à nos intuitions. Le principe de crédulité nous oblige à protéger nos intuitions les plus fortes en révisant les plus faibles. En conséquence, le meilleur compatibilisme est *a priori* celui qui va le plus loin dans la tentative d'accorder nos différentes intuitions. Fischer a raison de faire le maximum en ce sens, mais sa théorie est encore trop peu intuitive.

La responsabilité morale n'est pas liée à l'existence de dispositions mauvaises : on ne punit pas quelqu'un pour des dispositions ou des propriétés mauvaises. Au contraire, sauf dans les cas où il est responsable de ses dispositions, on excuse l'agent qui a des dispositions mauvaises, parce qu'il lui est difficile de bien agir. Elle n'est pas liée non plus, comme dans la théorie de Fischer, à l'existence d'une bonne disposition ou d'une bonne propriété. Faut-il alors penser qu'elle est liée uniquement aux caractéristiques de l'action elle-même,

³³ G. W. Leibniz, *Conversation sur la liberté et le destin*, 1703, p. 54.

indépendamment des dispositions ou des propriétés de l'agent ? Non plus. Cette hypothèse ne fonctionne pas mieux car l'action elle-même ne dit rien sur la responsabilité de l'agent : il peut ne pas être responsable (s'il agit sous l'effet d'une contrainte, etc.). Il faut donc que l'agent agisse par lui-même, mais non par une de ses propriétés, ni par une de ses dispositions. Or cette idée ne peut aboutir qu'à l'incompatibilisme.

Pour terminer ce chapitre sur les théories compatibilistes, je voudrai aborder le problème de manière plus frontale en posant la question suivante : quel est le type d'entité qui peut déterminer une action sans retirer à l'agent la responsabilité de cette action ? Si le déterminisme est vrai, la pertinence d'un monde possible pour mesurer la responsabilité d'un agent dépend de la nature des causes qui l'ont déterminé. C'est ce que nous avons pu voir dans les exemples de ce chapitre. Et pour un compatibiliste qui ne mesure pas la responsabilité de l'agent à partir de mondes possibles, il est également important de savoir quel genre de causes peut déterminer une action sans exclure la responsabilité, car on ne peut nier que certaines causes l'excluent. Y a-t-il donc une catégorie de causes ou de facteurs qui peuvent déterminer l'action d'une personne sans lui ôter sa responsabilité morale ? Il semble difficile d'en trouver une. Dans l'hypothèse où le comportement de l'agent est entièrement déterminé par le passé actuel et les lois de la nature, s'il n'agit pas de façon rationnelle dans le monde actuel, c'est qu'il en est empêché par un défaut, une cause interne, ou une cause externe. Si c'est un défaut qu'il l'en empêche, il n'est pas responsable : il lui manquait un moyen nécessaire, ou il y avait en lui un handicap ou un défaut de constitution. Si c'est une cause externe qui l'empêche d'agir mieux, on s'accorde en général pour dire qu'il n'est pas responsable. Mais si cette cause est interne à l'agent, la conclusion ne semble pas différente : par exemple, s'il a une maladie ou un parasite qui l'empêche d'agir autrement. Pour échapper à l'objection, il faut imaginer une distinction entre deux sortes de causes internes : celles qui excluent la responsabilité morale, et celles qui ne l'excluent pas. Mais alors quelles sont les causes internes qui empêchent l'agent d'avoir un comportement rationnel et qui diffèrent si radicalement des parasites ou des handicaps ? Je ne vois pas comment un compatibiliste peut répondre à cette question. De plus, par hypothèse, les causes internes à l'agent sont entièrement l'effet de causes externes. La distinction entre causes internes et externes ne paraît donc pas pertinente. Peut-on la

remplacer par une distinction entre deux genres de causes externes : celles qui agissent à travers la rationalité de l'agent, et celles qui agissent directement sur sa psychologie ou sur ses nerfs ? Je vais aborder maintenant ces deux distinctions. Je soutiendrai qu'aucune n'est utile pour concilier le déterminisme avec la responsabilité morale.

6.8 Causes externes ou internes, causes naturelles ou non-naturelles

L'hypothèse déterministe implique que dans une situation donnée, un agent ne peut agir de plusieurs façons différentes, à cause du passé actuel et des lois de la nature. Lorsque toutes les conditions suffisantes d'un événement sont réunies, il se produit inévitablement. Par exemple, il suffit de plonger un sucre dans l'eau pendant deux ou trois minutes pour qu'il se dissolve ; si ces conditions sont réunies, il se dissout inévitablement. De la même façon, le déterminisme implique que si toutes les conditions sont réunies pour qu'une personne agisse de telle ou telle manière, elle agit inévitablement ainsi. On en déduit que si un individu n'agit pas de façon rationnelle, c'est parce qu'il y a une *cause* ou un *défaut* qui le rend *incapable* d'agir de façon rationnelle. Par « défaut », entendons ici l'absence d'un moyen nécessaire. Ce moyen peut être immanent ou extérieur à l'agent. Dans le premier cas, c'est une faculté ou une propriété de l'agent ; dans le second, un outil, un instrument, ou une information que l'agent ne possède pas.

Il est clair que si l'action irrationnelle est due simplement à un défaut, il n'est pas responsable. Beaucoup de causes externes semblent également exclure la responsabilité de l'agent si elles déterminent son action. On pourrait donc croire qu'il est responsable d'une action irrationnelle seulement si une cause *interne* d'un certain genre a causé l'action. Mais lorsqu'une propriété de l'agent ou un événement survenant en lui l'empêche de réaliser une action bonne et provoque au contraire une action mauvaise, nous pensons normalement que cette propriété ou cet événement rend l'agent *incapable* d'agir autrement. On entend parfois dire qu'un homme est libre quand ses actions résultent de causes internes, c'est-à-dire de propriétés qu'il possède ou de faits survenants en lui. Mais cela ne s'accorde pas avec notre habitude d'excuser l'agent lorsque que son comportement est dû à une maladie ou à un défaut de constitution, quelle que soit la nature d'un tel défaut. Par exemple, nul n'est pas responsable

de ses crises d'allergie ou d'épilepsie, à moins de s'être lui-même placé dans des conditions où elles devenaient probables, et si l'agent est trisomique ou hydrocéphale, nous l'excusons plus volontiers que s'il est parfaitement sain. Or la trisomie est un défaut inscrit jusque dans les propriétés génétiques de l'individu ; loin d'être un défaut superficiel, elle est en relation avec l'identité profonde de l'individu et atteint à la fois le corps et l'esprit.

Faut-il distinguer deux types de causes internes ? On voit mal ce qu'une telle distinction apporterait en faveur du déterminisme, car si le déterminisme est vrai, toute propriété de l'agent qui l'empêche d'agir ou de vouloir rationnellement doit être vue comme un défaut dans sa constitution, un défaut qui le rend proprement incapable d'agir mieux. Wolf, Fischer et Ravizza, n'ont pas donné de solution à ce problème. Pour l'éviter, il faut tourner le dos à l'idée que les défauts de ce genre excluent ou limitent la responsabilité morale ; il faut donc adopter une conception de la responsabilité diamétralement opposée à celles de Wolf, Fischer et Ravizza, et encore plus éloignée du sens commun, celle que j'ai évoquée en citant un film de Fritz Lang.

A la fin de leur étude sur la responsabilité, Fischer et Ravizza admettent que leur théorie est en difficulté si on imagine une situation où l'agent est sensible aux raisons grâce à une manipulation électronique qui ne modifie pas le cerveau de l'agent³⁴. Rappelons en effet que pour eux, l'agent est autonome et responsable si ses actions sont causées par un mécanisme qui fait partie de lui (son cerveau, en l'occurrence) et qui est sensible aux raisons. En imaginant un savant fou qui manipule ainsi l'agent et augmente artificiellement sa sensibilité aux raisons, on devrait donc dire que cet agent est responsable de ses actes. Pourtant, son comportement est le résultat d'une manipulation opérée par un agent extérieur.

Le déterminisme implique que le mécanisme permettant aux hommes d'être sensible aux raisons est toujours le résultat de causes extérieures, puisqu'ils ne l'ont pas créé eux-mêmes. Personne ne s'est créé lui-même, et tout être dont l'existence a un commencement est le résultat de causes extérieures. On me répondra que dans l'exemple du savant fou, le mécanisme sensible aux raisons ne fonctionne pas de manière *naturelle*, puisqu'il est

³⁴ J. M. Fischer & M. Ravizza, *Responsibility and Control*, p. 235-236.

influencé par un procédé *artificiel*. Normalement, il est fabriqué lentement par un processus de division cellulaire où interviennent différents facteurs biologiques, puis il fonctionne en interaction avec le monde extérieur et la communauté humaine à laquelle appartient l'agent. Mais cette distinction entre des causes naturelles et non-naturelles répond-elle vraiment à l'objection ? Le point essentiel est que dans les deux cas, si le déterminisme est vrai, l'individu n'est pas responsable des ressorts de son action, et du coup, il n'est pas responsable de son action elle-même.

Cela apparaît plus clairement dans le cas inverse :

[Manipulation électronique] Un savant fou manipule Jean à son insu par un procédé électronique qui diminue sa sensibilité aux raisons sans modifier son cerveau. Après l'opération, Jean vole un disque, et il n'a aucun remord lorsqu'on lui reproche cette action. L'opération du cerveau est la cause de ce défaut.

Nul ne songerait à dire que Jean est réellement coupable. Par conséquent, il serait également étrange de soutenir que dans une situation inverse, Jean mérite des éloges alors que son comportement généreux s'explique par une opération analogue. De la même façon, si chaque homme est non seulement conditionné, mais déterminé par son éducation, même selon un processus parfaitement naturel, il n'est pas moralement responsable. Nous estimons habituellement que meilleure a été l'éducation d'un individu, moins grand est son mérite. C'est là une manière de penser dont on ne saurait se défaire que très difficilement. Ainsi, ni la distinction entre les causes naturelles et non-naturelles, ni la distinction entre les causes internes et les causes externes, ne semblent utiles pour délimiter les cas de responsabilité morale. Pour finir, nous pouvons envisager une troisième distinction. Une idée classique en philosophie est que nous sommes libres lorsque nous sommes déterminés par des *raisons* au lieu d'être déterminés par des *passions* ou par des *causes*. C'est le déterminisme des raisons.

6.9 Causes et raisons. Le déterminisme des raisons

Certains philosophes pensent que la liberté humaine est incompatible avec une nécessité de type causal ou avec l'action de certaines causes, mais non avec la nécessité des raisons.

C'est le cas de Leibniz. Il déclare que l'homme est libre quand il n'est pas déterminé par des passions :

Nous nous déterminons nous-mêmes < , et sommes libres > en tant que nous agissons, et nous sommes déterminés par dehors < et comme assujettis > en tant que nous pâtissons (...) il faut dire que nous sommes libres en tant que nous [suivons] < sommes déterminés à suivre > la perfection de notre nature, c'est-à-dire la raison ; mais que nous sommes esclaves en tant que nous suivons les passions < et les coutumes > ou les impulsions indélibérées que la raison n'a point formées auparavant à une habitude de bien faire³⁵.

D'après Leibniz, Dieu, en tant qu'être parfaitement rationnel, est entièrement déterminé par les raisons :

< l'indétermination en elle-même est > une imperfection où il ne faut point prétendre. Dieu même, qui est le plus parfait, est aussi le plus déterminé à agir de la manière conforme à la suprême raison, et l'indétermination apparente, < ou égalité en balançant > où nous sommes quelquefois vient de l'ignorance ou de l'impuissance (...).³⁶

Dieu est à la fois entièrement libre (car indépendant) et entièrement déterminé par les raisons. Il est assez naturel de penser que plus un être est libre, plus il agit de façon rationnelle. Aussi, puisque la liberté de Dieu paraît compatible avec une forme de nécessité qui détermine tous ses choix (la nécessité *des raisons*) la liberté humaine semble également compatible avec cette forme de nécessité. Cependant, l'analogie est contestable, même si elle dit quelque chose de vrai sur la nature de la liberté. Sans doute, un individu est autonome quand il se détermine lui-même selon les raisons. Mais il y a plusieurs manières de comprendre cette idée. Pour un déterministe, cela revient à dire qu'il est autonome quand il est déterminé par les raisons. Mais il faut bien voir à quel point le cas de l'homme diffère de celui de Dieu. Les raisons qui déterminent le comportement d'un homme viennent d'abord de son éducation, de la perception, de la mode et de l'opinion, donc de causes extérieures agissant sur l'individu. C'est en tout cas ce qu'implique la thèse déterministe, selon laquelle tout ce qui arrive est une conséquence nécessaire du passé actuel.

³⁵ G. W. Leibniz, *Conversation sur la liberté et le destin*, p. 50-51.

³⁶ *Ibid.*, p. 50.

Appelons C le choix de l'agent (choix qu'il fait librement, et dont il est responsable). Pour un déterministe des raisons, C est déterminé par des raisons de l'agent, qui se ramènent à des croyances et à des désirs de l'agent. Or ces croyances et ces désirs ont leur origine en dehors de l'agent, dans des causes qu'il ne contrôle pas et dont il subit l'influence. On peut difficilement nier l'influence de notre entourage sur nos croyances et nos désirs. A cela s'ajoute celle des choses que nous percevons et qui provoquent des désirs. Devant ce constat, un libertarien peut dire que nous sommes influencés sans être déterminés, mais un déterministe sera contraint de raisonner autrement. Appelons E un événement que l'agent ne contrôle pas. Le déterministe doit raisonner ainsi : Si E influence C (le choix), alors soit E détermine C, soit c'est un ensemble de causes incluant E qui détermine C. Par exemple, si Pierre croit, sous l'influence de sa grand-mère, que les feux follets sont des esprits, soit celle-ci détermine C (elle en est la cause suffisante), soit c'est un ensemble de causes qui détermine C, par exemple les paroles de la grand-mère, mais aussi la naïveté de Pierre ou son ignorance. Et on ne voit pas comment ce déterminisme pourrait ne pas être un déterminisme causal.

Il semble donc que si le déterminisme est vrai, la nécessité des raisons est chez l'homme une forme de nécessité causale. Elle ne l'est pas dans le cas de Dieu, parce que ses connaissances ne viennent pas de causes extérieures à lui. Mais le cas de l'homme est bien différent. Pour un incompatibiliste, les croyances d'un homme sont influencées, mais pas déterminées causalement. L'agent a la possibilité, dans une certaine mesure, de résister à cette influence. Pour un déterministe, les croyances de l'homme sont causées par le monde extérieur de façon nécessaire. Dans cette perspective, on ne voit pas comment l'homme peut être autonome en suivant les raisons.

De plus, le déterminisme des raisons se heurte à une difficulté concernant les actions irrationnelles dont l'agent est considéré comme responsable. En effet, on ne peut pas expliquer une faute ou un crime à partir des seules raisons, puisqu'un crime, par hypothèse, est un acte en partie irrationnel. Si le criminel est coupable, c'est que son acte n'a pas été entièrement déterminé par des causes, ni par des raisons. Mais alors, qu'est-ce qui a pu le déterminer ? Si l'acte libre - mais irrationnel - découle inévitablement du passé actuel, alors il est déterminé par des causes, et non seulement par des raisons ; car s'il n'était déterminé que par des raisons, il serait rationnel. Il faut donc admettre soit que les actions libres et irrationnelles sont

indéterminées, soit qu'elles sont déterminées par des causes, mais il n'y a pas de milieu entre ces deux hypothèses.

Ainsi, la distinction entre la nécessité des causes et celles des raisons ne permet pas d'expliquer la notion de responsabilité morale. Pour l'homme, être déterminé par des raisons, c'est être déterminé par des causes. C'est pourquoi il faut admettre que si les actions humaines sont nécessaires, elles le sont par l'effet de causes extérieures. Si les choix et les actions de l'agent sont causés par des raisons qui sont elles-mêmes causées par des événements ou des états de choses échappant au contrôle de l'agent, nous sommes en droit, une fois de plus, d'appliquer le principe du transfert et de conclure qu'il n'est pas responsable.

Résumé. Après avoir examiné (dans le chapitre IV) la solution compatibiliste fondée sur l'idée d'une hiérarchie de désirs, j'ai soutenu que face aux cas de responsabilité dans les omissions, les meilleures solutions compatibilistes sont celles qui attribuent à l'agent (dans les cas de responsabilité morale) un pouvoir des alternatives, et qui l'analysent en termes de mondes possibles. Elles suppose, comme la théorie de Fischer et Ravizza, que l'agent est responsable d'un mal lorsqu'on peut lui attribuer la faculté de l'éviter dans des situations voisines de la situation actuelle. Mais contrairement à ce que soutient Fischer, cette idée pose un problème de délimitation des mondes possibles pertinents. Puisqu'il y a une limite nette entre la responsabilité et la non-responsabilité, il faut montrer une limite nette entre les mondes possibles pertinents et les autres. Pour cela, le compatibiliste pourrait distinguer deux catégories de causes : les causes naturelles et non-naturelles, internes et externes, ou encore rationnelles et non-rationnelles. Or aucune de ces distinction ne semble pertinente. Même la troisième aboutit à de sérieuses difficultés. Premièrement, du point de vue déterministe, être déterminé par des raisons revient à être déterminé par des causes, les raisons ayant leur origine ultime en dehors de l'agent. Deuxièmement, si toute action libre est déterminée par des raisons, alors il est impossible qu'une action soit à la fois libre et irrationnelle, autrement dit, on ne serait jamais tout à fait responsable d'une action irrationnelle. Pour que l'influence des raisons sur l'agent ne le prive pas de toute responsabilité, il paraît donc essentiel que celui-ci

ait le pouvoir de les accepter ou de les refuser, ce qui nous ramène encore à l'idée de possibilités alternatives.

6.10 Conclusion. La balance des intuitions

Pour montrer la compatibilité du déterminisme et de la responsabilité morale, les théories compatibilistes essaient de montrer comment un agent peut exercer un certain contrôle de lui-même et de ses actions tout en étant déterminé causalement. Je pense que pour être acceptable, une théorie compatibiliste doit accorder une place non négligeable aux trois intuitions que j'ai mentionnées au début du premier chapitre, à savoir que l'agent est responsable de son action seulement s'il l'accomplit de manière volontaire, de manière autonome, et en ayant eu le pouvoir de ne pas l'accomplir.³⁷

L'analyse conditionnelle de la liberté (S est libre de faire $A \leftrightarrow$ Si S veut faire A , il le fait), avec quelques compléments, permet de satisfaire ces trois conditions. Elle ne suffit pas en-elle même, car seule, elle néglige le fait que si la volonté de l'agent est manipulée, il n'est pas vraiment libre. Il faut donc la compléter en l'appliquant au vouloir ou aux désirs. Cela introduit l'idée que l'agent contrôle ses propres désirs en les voulant ou en les désirant. Frankfurt, Watson, Taylor, et Lehrer, ont défini la liberté en élaborant une théorie hiérarchique des désirs fondée sur cette idée. Or d'un point de vue compatibiliste, elle aboutit à une difficulté. Si les désirs du niveau le plus élevé ne sont pas voulus eux-mêmes par l'agent, il ne les contrôle pas. Pour éviter que cela nous entraîne dans une régression à l'infini, il faut s'arrêter à des désirs qui ne sont pas eux-mêmes causés ou contrôlés par d'autres désirs. Chez Watson, certains désirs sont premiers et constituent le moi profond de l'agent ; et chez Lehrer, il n'y a pas non plus de cercle ou de régression à l'infini, car les préférences et l'ultrapréférence viennent ultimement des désirs de premier niveau que l'agent possède dès le début de sa vie. Finalement, cette idée conduit à penser que la liberté est liée à l'accord des actes et des désirs avec certaines valeurs fondamentales constituant le *moi profond* de l'agent.

³⁷ Cf. *supra* : 1.1 et 6.1.

Or on peut toujours imaginer un individu dont les valeurs profondes sont le résultat d'une manipulation.

D'un point de vue libertarien, il est plus facile d'éviter ce problème. On peut d'abord s'appuyer sur l'idée d'une connaissance pratique liée à l'autonomie ultime de l'agent : lorsqu'il agit librement, il sait ce qu'il fait et ce qu'il veut, parce qu'il est la source *ultime* de son action et de son vouloir. Son autonomie implique donc une conscience infiniment réflexive de son désir. En voulant librement, il sait ce qu'il veut, il sait qu'il veut ce vouloir, et ainsi de suite. Ensuite, la responsabilité de l'agent dans ses actions n'est pas liée à une certaine correspondance entre ses actions et un noyau de propriétés ou d'inclinations constituant le moi profond de l'agent. En effet, le libertarien peut définir l'identité profonde de l'agent à partir de l'autonomie ultime, en disant que l'agent est d'abord ce qu'il maîtrise en lui le plus directement.

Au lieu de définir la liberté comme le fait de faire ce qu'on veut et de vouloir ce qu'on veut, le compatibiliste aura donc intérêt à définir le contrôle de soi en lien avec la faculté d'agir rationnellement. C'est ce que tentent Wolf, Fischer et Ravizza. Cette approche a encore un avantage : elle permet d'expliquer la responsabilité morale dans les cas d'omissions. Lorsqu'un individu est responsable d'un événement par simple omission, cela ne vient évidemment pas du fait qu'il désire cet événement. Ici, la responsabilité ne se comprend pas à partir de l'attitude mentale de l'individu, ni de son action, puisqu'il n'agit pas. Elle ne peut se comprendre qu'à partir de ses facultés, propriétés, ou dispositions. Cela conduit nécessairement à une théorie comme celle de Fischer, fondée sur la sensibilité aux raisons. Mais cette dernière hypothèse demeure contraire à nos intuitions sur la responsabilité morale.

Dans le chapitre V, nous avons vu que les omissions posent un problème particulièrement délicat aux théories compatibilistes. En effet, lorsqu'on omet de s'opposer à un événement malheureux, on est responsable seulement si on a la possibilité d'empêcher cet événement. C'est ce que suggèrent tous les exemples du même genre que « Requins », où Jean ne peut sauver l'enfant de la noyade parce que des requins l'empêchent de s'en approcher. Ainsi, même si le principe PAP n'était pas correct, un principe analogue s'imposerait dans les cas d'omissions :

(PPA) (principe d'action possible) : un agent S est responsable par omission d'un événement E seulement s'il a eu la possibilité de faire une action A qui aurait empêché E de se produire.

Face à cette conclusion, le compatibilisme a le choix entre trois réponses. La première, fondée sur la réduction des omissions aux actions, est difficilement acceptable parce qu'elle néglige les asymétries qui apparaissent de manière assez claire entre les cas d'actions et les cas d'omissions. La deuxième et la troisième réponses ne nient pas PPA mais l'interprètent de manière compatibiliste. La deuxième réponse est l'analyse conditionnelle du mot « pouvoir » (ou « possibilité ») dans le principe PPA, et la troisième réponse analyse les pouvoirs de l'agent en termes de mondes possibles. La plus satisfaisante est la troisième réponse, mais elle se heurte pourtant à une série d'objections qui invalident aussi la deuxième réponse compatibiliste.

J'ai présenté ces objections dans le chapitre VI. La première n'est sans doute pas décisive, mais le faisceau des quatre objections suivantes justifie un sérieux doute sur la possibilité d'énoncer les conditions de la responsabilité morale d'une manière claire, cohérente, et compatibiliste. Tout d'abord, on ne peut pas définir les pouvoirs liés à la responsabilité en termes de *mondes possibles* sans délimiter l'ensemble des mondes possibles pertinents du point de vue de la responsabilité morale. Or rien ne permet de faire cela. Une définition des conditions de la responsabilité en termes de mondes possibles ne peut donc pas satisfaire l'exigence minimale de *clarté*. Deuxièmement, comme l'a montré Pereboom, on peut trouver des exemples où l'agent satisfait ces conditions alors qu'il est manipulé ou influencé dans le monde actuel par des facteurs causaux qui lui enlèvent sa responsabilité. Troisièmement, il y a aussi des cas où l'agent satisfait ces conditions sans être responsable de ce qu'il fait, parce qu'il lui manque un moyen nécessaire pour agir autrement.

Quatrièmement, la sensibilité modérée aux raisons n'est pas vraiment une disposition, sinon nous pourrions l'attribuer sans accorder une importance particulière aux mondes possibles les plus proches du monde actuel. En fait, le concept de sensibilité aux raisons élaboré par Fischer et Ravizza est un hybride entre une disposition et le pouvoir des alternatives compris de manière libertarienne. Le résultat de ce mélange est qu'il n'y a pas de

limite précise entre la sensibilité modérée et la sensibilité faible aux raisons, alors qu'il y en a une entre la responsabilité et l'absence de responsabilité. Puisque la sensibilité aux raisons, d'un point de vue compatibiliste, ne peut se définir en termes de possibilités ouvertes à l'agent *dans le monde actuel*, elle se réduit à une faculté d'agir mieux dans des situations voisines du monde actuel, sans que l'on puisse préciser lesquelles. Cela veut dire qu'un agent est responsable d'une mauvaise action lorsqu'on peut dire qu'il agirait mieux dans des situations voisines du monde actuel. Mais à la réflexion, une telle idée est étrange. En effet, prouve-t-on la culpabilité de quelqu'un en montrant qu'il agirait mieux dans des situations différentes ? On ne peut reprocher à l'agent d'avoir des dispositions qui le font agir mieux dans d'autres circonstances. S'il n'avait pas ces dispositions, il serait moins bon. On peut évidemment reprocher à quelqu'un de ne pas avoir telle ou telle bonne disposition. On peut également lui reprocher de ne pas avoir agi mieux s'il pouvait agir mieux *dans le monde actuel*, mais il est incohérent de mélanger ces deux intuitions.

Finalement, il semble plus cohérent d'opter pour l'idée inverse en supposant que les défauts qui empêchent l'agent d'agir mieux *fondent* la culpabilité au lieu de l'exclure. Mais cette idée s'oppose à notre habitude d'excuser l'agent lorsqu'un handicap ou une propriété quelconque l'empêche d'agir autrement. Elle suppose au contraire que l'agent est responsable des actions (ou des événements) qui découlent de ses propriétés, mais cette hypothèse a des conséquences trop absurdes, par exemple, l'agent serait responsable de ses crises d'allergie et de toutes les réactions de son organisme. Pour éviter cette conséquence, il faudrait distinguer deux classes de propriétés de l'agent, en indiquant laquelle est liée à son identité profonde et à sa responsabilité morale. Mais on ne voit pas quelle pourrait être cette distinction.

De plus, selon cette théorie, il suffirait de connaître les propriétés d'un individu pour savoir s'il est coupable. Il serait donc inutile d'en connaître l'histoire, idée contre-intuitive, là aussi. Comme le soulignent Fischer et Ravizza, la notion de responsabilité morale est une notion *historique*. Mais les libertariens sont mieux placés pour rendre compte de ce fait. Car si la responsabilité de l'agent implique qu'il ne soit pas déterminé causalement, il ne suffit pas de connaître ses propriétés pour savoir s'il est responsable, mais il faut savoir dans quelle mesure il est la source *ultime* de ses propriétés, ce qui peut nous faire remonter très loin dans l'histoire personnelle de l'agent.

Pour répondre à ces objections, le compatibilisme a besoin de classer les propriétés et dispositions de l'agent en deux catégories : dans l'une, celles qui viennent d'ailleurs et ne le constituent pas vraiment ; dans l'autre, celles qui constituent l'identité profonde de l'agent même si elles viennent aussi de causes extérieures. A partir de là, on pourrait distinguer deux types de situations : celles où l'agent est responsable d'un événement, et celles où il cause un événement sans en être responsable. Mais pour rendre compte du caractère historique de la notion de responsabilité, il faut supposer, dans cette perspective, que les propriétés profondes de l'agent viennent en premier d'un point de vue chronologique. Il est alors utile d'évoquer l'enfance d'un criminel pour savoir si ses dispositions mauvaises appartiennent à son moi profond, c'est-à-dire originel. Si elles n'existent pas dans son moi originel, on déclare qu'elles viennent d'influences plus ou moins tardives et que l'individu n'en est pas responsable. Néanmoins, cette distinction purement chronologique n'est guère satisfaisante. Nous ne considérons jamais un individu comme responsable des propriétés qu'il a en naissant. Au contraire, nous le considérons comme responsable de propriétés qu'il acquiert plus tard. Un libertarien est mieux placé pour expliquer cet état de choses : il peut dire que l'individu n'est jamais la *source ultime* des propriétés qu'il a en naissant, alors qu'il est à l'origine de certaines propriétés acquises plus tard au cours de son développement.

Comme le soulignent les compatibilistes, il est possible de louer et d'aimer une personne pour des qualités venant de causes qu'elle ne contrôle pas. On peut louer, par exemple, la beauté d'une personne. Cependant, la louange et le blâme sont-ils de cet ordre dans le domaine moral ? Ce que nous avons vu implique que les jugements moraux portant sur la responsabilité des personnes ne sont pas du même ordre, mais supposent en elles une faculté d'agir sans y être déterminé causalement.

Enfin, toutes les théories déterministes ou compatibilistes se heurtent à une autre intuition très forte, que je ne ferai qu'évoquer ici. Dans notre vie la plus quotidienne, nous sommes intimement persuadés qu'il y a des futurs contingents, que certaines possibilités nous sont ouvertes. Si tous nos choix découlent inévitablement du passé actuel et des lois de la nature, il y a une foule de comportements, d'expressions et d'usages linguistiques dont la

signification devient énigmatique, quand elle ne disparaît pas purement et simplement. En effet, que deviennent alors promesses, conseils, menaces, encouragements, exhortations et interdictions ?

Arrêtons-nous un moment sur le cas des promesses. Que peuvent-elles exprimer si nos choix sont entièrement prédéterminés ? Lorsqu'on promet d'agir bien, on exprime seulement une certitude sur son propre comportement futur ? Par exemple, si je promets à un ami de lui rendre un livre avant une certaine date, ma promesse signifie-t-elle seulement qu'à mon avis, je lui rendrai ce livre avant la date convenue, bien que cela dépende entièrement de causes que je ne contrôle pas ? Cette réponse est peu intuitive et ne montre pas l'utilité des promesses. D'un point de vue déterministe, la promesse n'est utile que si elle *détermine* elle-même le comportement futur de l'agent dans une bonne direction. Si ce n'est pas la promesse *elle-même* (ajoutée à d'autres causes) qui détermine le comportement futur de l'agent dans le sens de ce qui est promis, elle ne semble d'aucune utilité.

Il y a sans doute des intuitions dont les théories compatibilistes peuvent rendre raison, mais elles n'y arrivent souvent qu'en les vidant d'une partie de leur contenu. L'idée d'autonomie, par exemple, a un sens beaucoup moins fort dans l'hypothèse déterministe que dans l'hypothèse contraire, où l'agent est partiellement indéterminé. Pour savoir quelle est la meilleure théorie, il faut savoir quelle théorie parvient le mieux à rendre raison de nos intuitions, et après un examen des voies empruntées par les compatibilistes, nous constatons qu'elles ne sauvent pas suffisamment nos intuitions sur la responsabilité. Par exemple, la théorie de Fischer et Ravizza propose une explication du lien entre le contrôle de soi et la raison, mais un examen attentif révèle que leurs définitions ont des conséquences contraires à nos intuitions.

La question qui se pose alors est de savoir s'il faut conserver le concept de responsabilité morale que nous héritons des siècles passés, ou l'abandonner parce que le monde physique nous apparaît comme entièrement soumis au déterminisme. Le second choix est celui des incompatibilistes « durs ». Pereboom³⁸ estime que nous ne pouvons pas

³⁸ Cf. Derk Pereboom, *Living without Free Will*, Cambridge University Press, Cambridge, 2001.

maintenir la conception traditionnelle de la responsabilité ; le plus important à ses yeux est de justifier les pratiques juridiques auxquelles nous ne pouvons pas renoncer. La difficulté est qu'elles se fondent traditionnellement sur la notion métaphysique du libre arbitre que rejette précisément Pereboom. Une tâche essentielle pour lui est donc de proposer une nouvelle justification de ces pratiques. Pourquoi emprisonner les criminels s'ils ne sont pas responsables ? Le but des châtiments est-il de dissuader ? Dans ce cas, pourquoi hésiter à infliger des peines plus sévères ? Elles seraient plus efficaces... Ou bien l'emprisonnement est-il une sorte de quarantaine destinée à préserver la société d'un mal contagieux ?

Nous ne suivrons pas Pereboom dans le traitement de ces questions. Avant d'abandonner le concept de responsabilité morale, il faut savoir s'il y a une raison impérieuse de le faire. Si aucune raison ne nous oblige à adhérer au déterminisme, rien ne nous oblige à changer nos croyances sur la liberté humaine. Le compatibilisme ne semble pas pouvoir résoudre les difficultés que nous avons abordées dans les chapitres V et VI. Il ne s'imposera donc comme solution que si des raisons suffisamment fortes nous obligent à adhérer au déterminisme. Dans le chapitre VII, je soutiendrai qu'il n'y a pas d'argument décisif en faveur du déterminisme. Mais l'existence de phénomènes indéterminés ne suffit pas pour que le libre arbitre puisse s'insérer dans la causalité physique. L'indéterminisme à l'intérieur de l'agent peut même s'opposer au contrôle de soi. Je m'efforcerai donc de montrer à quelles conditions il peut favoriser le contrôle de soi et la responsabilité morale. Enfin, j'essaierai d'indiquer où sont les questions les plus difficiles à résoudre du point de vue libertarien.

VII Le libre arbitre et le monde physique

Déterminisme et indéterminisme

7.1 Introduction

Lorsqu'une affirmation non évidente ne peut pas se démontrer de manière rigoureusement déductive, sa probabilité s'évalue par comparaison avec la probabilité de l'affirmation contraire. Ainsi, pour savoir si la thèse libertarienne est plus probable que la thèse déterministe, il faut comparer la force des arguments libertariens avec celle des arguments déterministes ou anti-libertariens. Pour cette raison, il est intéressant d'examiner non seulement les *théories compatibilistes*, comme je l'ai fait dans les chapitres IV à VI, mais aussi les *arguments déterministes*. J'ai soutenu qu'aucune théorie compatibiliste ne peut rendre raison d'une manière cohérente de toutes nos intuitions sur la responsabilité morale, mais s'il y avait des arguments de poids en faveur du déterminisme ou contre l'idée d'indéterminisme, il se pourrait qu'une théorie compatibiliste, malgré ses incohérences ou son aspect contre-intuitif, s'impose en tant que théorie *la moins incohérente* ou la moins contre-intuitive. Vu les difficultés rencontrées par la thèse compatibiliste, il faut avoir de solides raisons pour la préférer à la thèse libertarienne. Dans ce dernier chapitre, je vais donc aborder les objections contre la thèse libertarienne. Tout d'abord, j'exposerai un argument pour montrer qu'il n'y a aucune preuve solide en faveur de la thèse déterministe. Ensuite, à partir de la section 7.5, je discuterai les deux grandes théories libertariennes : celle de la causalité événementielle (Kane, Ekström), puis celle de la causalité de l'agent (Chisholm, Taylor, O'Connor). La thèse libertarienne implique qu'il y a de l'*indéterminisme* à la source de nos actions. Or un événement indéterminé relève normalement du hasard ou échappe à notre contrôle. Ce ne sont donc pas seulement les déterminismes qui s'opposent à la liberté. L'existence de processus *indéterminés* à la source des actions paraît limiter aussi le contrôle de soi et la responsabilité morale. Cet argument appelé « *Mind argument* » se présente sous diverses formes. Je l'exposerai (en 7.8) sous une première forme, qui s'applique indifféremment aux deux grandes théories libertariennes. Ensuite, j'exposerai une deuxième version de l'argument qui met en difficulté la théorie de la causalité événementielle. Enfin, en 7.11 et 7.12, j'examinerai deux

objections contre la théorie de la causalité de l'agent. Je n'essaierai pas de répondre entièrement à ces objections, mais de montrer qu'elles ne nous obligent pas à abandonner la thèse libertarienne. En effet, cette thèse est cohérente et conforme à nos intuitions. Le fait que le libre arbitre ait quelque chose de mystérieux ou d'énigmatique ne veut pas dire qu'il faut cesser d'y croire. La vie et la pensée sont des réalités mystérieuses, mais personne ne doute de leur existence. De la même façon, il est possible que la liberté humaine soit une réalité mystérieuse ou peu intelligible.

7.2 « Tout a une cause », « rien ne vient du néant ». Le principe de causalité

Le déterminisme est la thèse selon laquelle tout événement est causé de façon nécessaire. Pour établir cette thèse, il faut donc prouver deux affirmations :

(a) Tout événement a une cause.

(b) Toute relation de cause à effet est une relation nécessaire, c'est-à-dire telle que si cette cause se produit, alors il est nécessaire que cet effet se produise.

De là, on peut déduire :

(c) Tout événement est causé de façon nécessaire. Autrement dit, tout événement est déterminé causalement.

Mais aucune des deux prémisses de l'argument n'est immédiatement évidente. Pour que la conclusion s'impose, il faut donc que ces deux affirmations soient prouvées de manière déductive, ou de manière inductive. Dans développements qui suivront, je soutiendrai que si la prémisse (a) est solide en un certain sens, la prémisse (b), en revanche, n'a pas de preuve déductive ni inductive, ce qui doit nous empêcher d'adhérer à la thèse déterministe.

Pour justifier la thèse déterministe, on invoque souvent le « principe de causalité ». Mais il n'est pas toujours facile de savoir quelle est sa signification. Tout est soumis, dit-on, à la causalité universelle, mais qu'entend-on par là exactement ?

Je pense qu'en invoquant ce principe, on se réfère à quelque chose comme :

(PC1) « Rien ne peut naître à partir du néant ». Ou bien :

(PC2) « Tout a une cause ».

La première formulation du principe de causalité (PC1) se trouve dans des textes très anciens, par exemple chez Epicure, dans la *Lettre à Hérodoté* :

Tout d'abord, rien ne naît de rien, autrement tout pourrait naître de tout sans avoir besoin d'aucune semence. Et si ce qui disparaît était réduit à rien, toutes choses auraient déjà péri, étant donné que celles en lesquelles elles se sont dissoutes n'existeraient pas.¹

On la rencontre aussi dans l'ancienne sagesse de l'Inde :

Comment l'être naîtrait-il du néant ? En vérité, c'est l'être qu'il y avait au commencement, l'être seul et sans second. (*Chandogya Upanishad*, VI, 2)²

Ce principe paraît solide, mais ne permet pas directement de soutenir une thèse déterministe. En revanche, PC2 semble fournir un argument déterministe, mais il n'est pas aussi évident que PC1. Il est donc intéressant de voir si PC2 est une conséquence de PC1.

PC2 est une formulation du principe de causalité aussi vieille que PC1, et se confond facilement avec lui. On le trouve par exemple chez les stoïciens. Plutarque écrit dans le *De fato* que « rien ne se produit sans cause »³, et selon Alexandre d'Aphrodise, s'il y avait quelque événement dû au hasard et privé de cause (*anaitiôs*), le cosmos perdrait son unité : « l'univers ne serait plus administré selon un ordre et un plan unique, si quelque événement sans cause s'y introduisait »⁴. Mais la différence entre PC1 et PC2 est évidente, car si rien n'est sans cause, il n'y a pas de cause première incausée, mais une série infinie de causes, infinie parce que sans commencement. PC1, quant à lui, n'aboutit pas à cette conclusion. Ce

¹ Epicure, *Lettre à Hérodoté*, 39, trad. Maurice Solovine, dans *Épicure, Doctrines et Maximes*, Hermann & C^{ie}, Paris, 1965. (p. 38 dans *Épicure et les épicuriens*, textes choisis par Jean Brun, PUF, Paris, 1961). La première partie de la phrase (« οὐδὲν γίγνεται ἐκ τοῦ μὴ ὄντος ») peut se traduire « rien ne naît du non-étant ».

² Trad. Émile Senart, Belles Lettres, Paris, 1900, p. 78.

³ Plutarque, *De fato*, 31, 201, 32, trad. Frédérique Ildefonse, dans F. Ildefonse, *Les Stoïciens*, vol.1, Belles Lettres, Paris, 2000, p. 40.

⁴ Alexandre d'Aphrodise, *De fato*, 31, 201, 32, trad. Victor Goldschmidt dans son livre intitulé *Le système stoïcien et l'idée de temps*, Vrin, Paris, 1977, p. 83.

principe (« rien ne vient du néant ») signifie d'abord que si, à un instant donné, il n'y a absolument rien, il ne peut y avoir quelque chose à un instant ultérieur. Cela impliquerait que le néant soit productif, qu'il puisse causer quelque chose. Or le néant n'est pas une chose. Ou bien cela supposerait des êtres capables de se donner à eux-mêmes l'existence ; mais pour cela, il faudrait qu'ils existent avant de commencer à exister, ce qui est contradictoire.

PC1 exclut toute génération spontanée. Supposons qu'une chose naisse sans être causée par quelque chose de préexistant. Cette hypothèse aboutit de toute façon à une absurdité : en effet, soit elle serait produite par le néant, soit elle se créerait elle-même. On peut donc conclure de PC1 que *tout ce qui a un commencement a une cause*. Or cela n'exclut pas l'existence de choses ou d'entités qui n'ont pas de commencement⁵. Par exemple, l'idée aristotélicienne d'une cause première ou d'un « Premier Moteur non-mû » n'implique pas de *génération* spontanée puisque cette cause est éternelle et sans commencement. Cette cause n'est pas un effet, ni d'autre chose, ni d'elle-même. Elle ne sort pas non plus du néant puisque son éternité même exclut que le néant ait existé dans le passé.

⁵ Lucrèce, comme d'autres après lui, tire de PC1 des conséquences qui n'en découlent pas de manière évidente : "Notre premier principe, c'est que rien n'est jamais créé de rien par l'effet d'une puissance divine. En effet, la crainte subjugué tellement les cœurs des mortels, qu'à la vue des phénomènes du ciel et de la terre, dont ils ne pouvaient pénétrer les causes, ils ont imaginé des dieux créateurs. Quand nous nous serons assurés que rien ne se fait de rien, nous distinguerons plus aisément le but où nous tendrons, la source d'où sortent les êtres, et la manière dont chaque chose peut se former sans le secours des dieux.

Si quelque chose s'engendrait de rien, les êtres de toute espèce pourraient naître indifféremment de toute sorte de corps, sans avoir besoin de germes particuliers. L'homme pourrait naître dans les ondes, les poissons et les oiseaux se former dans la terre, les troupeaux s'élancer des nues, et les bêtes féroces, enfants du hasard, se plaire également dans les lieux cultivés ou dans les déserts ; les arbres n'offriraient pas constamment les mêmes fruits : ils en changeraient chaque jour ; tous les corps pourraient produire les fruits de toute espèce. [...] les corps tirés du néant n'auraient pas besoin, pour croître, du temps et de la réunion de leurs germes ; l'enfant deviendrait tout à coup jeune homme, et l'arbuste à peine éclos s'élancerait vers la nue. Ce n'est pas là la marche de la nature : la fixité des éléments assujettit les corps à des progrès lents ; tout en croissant, ils conservent leur caractère, preuve évidente que chaque être a sa matière propre qui sert à le nourrir et à le développer." Lucrèce, *De Natura Rerum*, I, v.151 et s., trad. A. Ernout, Belles Lettres, Paris, 1924, traduction reprise par Jean Brun dans *Épicure et les épicuriens*, textes choisis par Jean Brun, PUF, Paris, 1961, p.70.

On observe ici un glissement d'une idée à une autre, voisine mais non équivalente. Tout d'abord, Lucrèce passe de « rien ne peut sortir du néant » à l'idée qu'on ne peut rien produire sans une matière appropriée. Ensuite, il déclare que si ce principe n'était pas vrai, il n'y aurait aucune régularité dans la nature (dans la succession des phénomènes). En effet, si quelque chose pouvait sortir du néant, les choses pourraient commencer à être sans que leur apparition requière certaines causes nécessaires ; elles pourraient surgir à tout moment et dans n'importe quelles conditions. Il n'y aurait donc aucune régularité dans la succession des événements. Mais, en admettant que PC1 soit en effet valable, il n'exclut pas qu'une chose soit produite sans matière préexistante. Il signifie seulement que rien ne peut être produit sans avoir de cause. Lucrèce s'attaque aux mythes racontant la création des choses par les dieux. Mais si Zeus produit un être sans employer aucune matière, on ne peut pas dire que cette créature sorte du néant, car il y avait quelque chose avant elle et qui lui donne l'existence.

Cela a une application dans le débat sur le libre arbitre. En effet, le libre arbitre suppose que l'homme soit la source ultime de certaines actions. Or l'existence d'un être humain a un début, contrairement à celle de Dieu. Pour que l'idée de libre arbitre soit recevable, il faut donc que les actions d'un homme aient leur origine en lui. Ainsi, elles ne viennent pas du néant. Mais cela ne fait que repousser le problème, car si l'agent doit causer ses mouvements sans être lui-même causé par rien, on se heurte apparemment au principe PC1 : ce qui n'a pas toujours existé a nécessairement une cause. Or comment l'agent humain peut-il être une sorte de « cause première » s'il n'a pas toujours existé ? La réponse est simple : chaque être humain est né à partir des générations antérieures (et a donc une cause), mais lorsqu'il agit librement, ses actions ne sont pas entièrement déterminées par des causes. L'existence du libre arbitre ne suppose pas que l'agent soit *causa sui* quant à son existence. Un homme agit librement lorsqu'on peut le considérer comme une source ultime de ses actions, mais son existence même a une origine causale hors de lui.

Chez Hume, nous trouvons le principe PC2 dans un passage bien connu où il soutient que la liberté n'implique pas la contingence :

Tout le monde reconnaît que rien n'existe sans une cause de son existence et que le mot hasard, quand on l'examine de près, est purement négatif et ne désigne aucun pouvoir réel qui existerait quelque part dans la nature. Mais on prétend que des causes sont nécessaires et d'autres non⁶.

En lisant Hume, on peut se demander à quoi il applique le principe selon lequel rien n'est sans cause : pense-t-il aux choix, aux actions ou à l'agent lui-même ? L'idée de possibilités alternatives n'est pas liée à celle d'un agent éternel comparable à Dieu, sinon les libertariens n'oseraient pas considérer l'homme comme ayant cette forme de liberté. En disant que le libre arbitre fait de l'homme une sorte de « mini-dieu », un « premier moteur non-mû », on veut seulement dire que ses choix et ses actions ne sont pas entièrement déterminés par des causes. Mais cela s'entend de plusieurs manières. Or les différentes solutions proposées par les libertariens actuels respectent le principe PC1. En effet, aucun libertarien ne prétend qu'il y ait des actions sans cause ou sans origine. Il dira plutôt que les actions accomplies librement sont causées par l'agent lui-même (que rien ne détermine à agir ainsi), ou bien qu'elles sont

⁶ D. Hume, *Enquête sur l'Entendement humain*, section VIII, "Liberté et nécessité", première partie, p. 164 dans la trad. de Philippe Baranger et Philippe Saltel (1947).

causées de façon contingente. La première idée a été défendue récemment par Taylor, Chisholm⁷ et O'Connor, et la seconde, par Kane⁸ et Ekström.

Dans l'exercice du libre arbitre, il n'y a donc aucun passage du non-être absolu à l'être, qui violerait le principe selon lequel rien ne vient du néant. La liberté implique seulement que même si le comportement de l'agent a plusieurs causes, le fait qu'il choisisse telle possibilité plutôt que telle autre n'a pas d'autre origine que *lui-même*. L'existence de l'individu a une cause extérieure à lui, le fait qu'il soit blanc ou noir, blond ou brun, en a une également, mais nous pensons qu'il y a des états ou propriétés qu'il possède par lui-même, au sens où il en est l'origine ultime. La thèse libertarienne ne viole donc pas le principe PC1, mais PC2.

Au lieu de se fonder sur PC2, qui paraît trop difficile à justifier, le déterminisme peut se fonder sur PC3, affirmation très proche de PC2 mais moins générale :

(PC3) : « tout événement a une cause ».

Puisqu'une action peut être vue comme un événement d'un certain type, on peut déduire de PC3 que toute action a une cause, y compris les actions libres. Contrairement à PC2, PC3 semble être une conséquence de PC1, car tout événement implique un changement ou une nouveauté, et une nouveauté qui n'a pas de cause viole PC1, puisqu'elle surgit à partir de rien. L'argument se présente ainsi :

- (1) Toute action est un événement.
- (2) Tout événement a une cause. (PC3)
- (3) Toute relation de cause à effet est une relation nécessaire.
- (4) Donc toute action est causée de manière nécessaire.

En admettant la prémisse (1), le libertarien peut répondre à l'argument en soutenant que la prémisse (3) est fausse. Pour cela, il doit montrer qu'il peut y avoir des relations causales

⁷ Roderick M. Chisholm, *Person and Object* (1976), ch. II. Ce chapitre expose une théorie de la causalité de l'agent (*agent-causal theory*). Kane, quant à lui, propose une théorie de la causalité événementielle (*event-causal theory*). cf. note suivante.

⁸ Robert Kane, *The Significance of Free Will* (1998), ch. VIII. Pour Kane, les décisions libres sont causées de façon indéterministe. Elles ont leur origines dans des événements physiques indéterminés à l'intérieur du cerveau. Ces événements sont causés, mais pas de façon déterministe. Nous reviendrons sur cette théorie dans la section 7.9.

indéterministes entre la cause et l'action qui en résulte. Admettons que l'agent appartienne à la catégorie ontologique des substances, ou à une autre catégorie que celle des événements. Dans cette hypothèse, puisque la prémisse (2) (« tout événement a une cause ») ne s'applique pas à l'agent lui-même (parce qu'il n'est pas un événement ni un ensemble d'événements), le déterministe ne pourra pas employer un argument comme celui que je viens d'exposer pour montrer que l'agent est déterminé par des causes antérieures. Et si nous voulons éviter le causalisme, nous pouvons soutenir que l'agent n'est pas au sens propre la *cause* de son action, comme si elle était extérieure à lui, mais sa *source* ou son *origine causale* en un sens plus large. La plausibilité de cette hypothèse n'est pas évidente, mais on ne peut l'écarter d'emblée. C'est pourquoi le déterminisme ne s'impose pas *a priori*. Dans les sections 7.3 et 7.4, je montrerai pourquoi la prémisse (3) est douteuse. J'en conclurai qu'il n'y a aucun argument solide en faveur de la thèse déterministe. Dans la suite du chapitre, je m'intéresserai aux autres objections contre la thèse libertarienne. Ces objections sont toutes fondées sur une opposition apparente entre l'indéterminisme et les notions de choix, de contrôle, et de rationalité.

7.3 Toute relation causale est-elle nomologique ?

Les choix que nous qualifions de « libres » sont souvent influencés par des causes. Ainsi en est-il de deux personnes faisant les courses ensemble : celle qui achète est influencée par le jugement de l'autre, et pourtant nous estimons qu'elle agit librement. Il arrive aussi que nos choix libres soient influencés par des sensations involontaires ou par des suggestions de notre imagination, que nous suivons spontanément ou contre lesquelles nous essayons de lutter. Si pendant mon travail, je suis saisi par la beauté de la lumière au moment où le soleil descend vers l'horizon, je peux avoir l'idée d'interrompre mon activité pour me promener dans la campagne. En supposant que j'avais la capacité de choisir autrement, la contingence de mon choix ne signifie pas qu'il soit sans cause extérieure. Puisqu'il est dû en partie à mon état psychologique au moment où j'ai admiré la lumière, on peut dire que j'ai été *influencé* par des sensations visuelles. Ainsi, bien que j'aie été influencé jusque dans mon choix, c'est un choix libre et contingent. Il semble donc que les libertariens doivent recourir à l'idée d'une action causale dont l'effet ne se produit pas nécessairement, à moins de considérer l'influence dont nous venons de parler comme étant de nature non causale, mais on voit mal comment une

influence peut exister entre deux choses si elle n'est pas une action causale, ou si elle ne s'accompagne pas d'une telle action.

Il semble donc que Hume ait bien posé le problème⁹ : le libre arbitre, tel qu'il est compris habituellement, suppose l'existence de relations causales où la cause ne produit pas nécessairement son effet, même si elle agit effectivement dans tous les cas. Cette idée est-elle acceptable ? Hume la rejette pour des raisons épistémologiques liées à la manière dont nous découvrons les relations de causalité. Pour lui, le sens même du mot « cause » implique une connexion nécessaire, car c'est en observant des connexions constantes entre les phénomènes que nous venons à l'idée que certains sont la cause des autres. Je vais soutenir maintenant que cet argument n'est pas concluant et que rien, dans le concept de cause, ne nous oblige à croire que toute relation de cause à effet soit une relation nécessaire. Il restera à examiner si cette idée peut se fonder sur une preuve *inductive*. Je montrerai dans la section suivante pourquoi, à mon avis, une telle preuve n'existe pas et ne *peut pas* exister.

Le déterminisme se fonde généralement sur l'idée que toute relation causale est *nécessaire* ou « *nomologique* ». Le mot grec « *nomos* » signifie « loi », mais l'adjectif « *nomologique* » est employé en général pour qualifier les relations causales nécessaires ; il ne sert pas dans le cas des relations dites « probabilistes ». Conformons-nous à cet usage, en laissant de côté la question de savoir si une loi est toujours strictement nécessaire. La conception humienne de la causalité a eu une grande influence, et correspond assez bien à notre représentation spontanée de la causalité. Selon Hume, le concept de cause est tiré de l'observation empirique d'une *connexion constante* entre deux types d'événements. Cela suggère que toute relation de cause à effet est *nécessaire*.

L'idée que toute relation causale est *nomologique* revient à dire que les mêmes causes produisent les mêmes effets. Sous cette forme, elle a un air de famille avec le principe d'identité des indiscernables, selon lequel si deux choses ont exactement les mêmes propriétés (ou si rien ne permet de les distinguer) on doit conclure qu'il s'agit d'une même chose. Le principe qui nous intéresse dit ceci : *une même chose, placée dans des situations identiques, doit avoir le même comportement, parce qu'elle a les mêmes propriétés*. Or c'est justement ce

⁹ Cf. David Hume, *Enquête sur l'Entendement humain* (1758), section VII, "L'idée de connexion nécessaire", première partie.

que nie la conception traditionnelle du libre arbitre comme capacité d'agir de différentes façons dans des situations identiques. Selon cette conception, la liberté de l'agent, dans une situation donnée, implique normalement qu'il soit capable d'agir autrement. Il y a donc plusieurs scénarios possibles, dans lesquels des actes différents sont causés par le même agent, doté pourtant des mêmes pouvoirs et des mêmes propriétés, et placé à chaque fois dans la même situation initiale. Au contraire, en disant que *les mêmes causes produisent nécessairement les mêmes effets*, on suppose que l'action est déterminée par un ensemble de causes. Aux yeux d'un déterministe, ce principe a sans doute l'apparence d'une vérité première, mais précisément parce que si le déterminisme est vrai, il en est une conséquence immédiate. Pour éviter un cercle vicieux, il faut trouver une raison indépendante d'accepter cette idée. Sur quel argument faut-il donc s'appuyer pour soutenir que toute relation causale est nomologique ?

Chez Hume, cette idée a un fondement épistémologique. Dans *l'Enquête sur l'entendement humain*, il insiste sur le caractère mystérieux des pouvoirs ou des énergies que nous attribuons aux causes. Ces pouvoirs ne sont pas des qualités sensibles, ni des entités que nous pourrions percevoir plus ou moins directement :

Quand nous regardons autour de nous les objets extérieurs et que nous considérons l'opération des causes, nous ne sommes jamais capables, dans un cas singulier, de découvrir un pouvoir ou une connexion nécessaire, une qualité qui lie l'effet à la cause, et fasse de l'un la conséquence infaillible de l'autre. Nous trouvons seulement que l'un, dans les faits, suit réellement l'autre. L'impulsion d'une des boules de billard est accompagnée du mouvement de la seconde boule. C'est tout ce qui apparaît aux sens *externes*. L'esprit ne sent aucun sentiment, aucune impression *interne* venant de la succession des objets. Par conséquent, rien, en un cas singulier, particulier, ne peut suggérer l'idée de pouvoir ou de connexion nécessaire.

À partir de la première apparition d'un objet, nous ne pouvons jamais conjecturer quel effet en résultera. Mais si le pouvoir ou l'énergie d'une cause pouvait être découvert par l'esprit, nous pourrions prévoir l'effet, même sans expérience, et nous pourrions, dès le début, nous prononcer avec certitude sur cet effet, par la seule force de la pensée et du raisonnement.¹⁰

¹⁰ David Hume, *Enquête sur l'Entendement humain*, trad. Philippe Foliot, GF-Flammarion, Paris, 2002, section VII, "L'idée de connexion nécessaire", première partie.

Hume suppose donc que l'existence des forces et des pouvoirs est inférée à partir d'observations répétées :

quand une espèce particulière d'événement a toujours, dans tous les cas, été jointe à une autre, nous ne faisons dès lors plus de scrupules à prédire l'un des événements sur l'apparition de l'autre, et à employer ce raisonnement qui peut seul nous assurer des choses de fait et d'existence. Nous appelons alors l'un la *Cause*; l'autre l'*Effet*. Nous supposons qu'il y a une connexion entre eux, un pouvoir dans l'un, par lequel il produit infailliblement l'autre, et qui opère avec la plus grande certitude et la plus forte nécessité.

Il apparaît donc que cette idée de connexion nécessaire entre des événements naît d'un certain nombre de cas semblables qui se produisent et de la constante conjonction de ces événements.¹¹

Pour Hume, la croyance qu'une chose est la cause d'une autre repose sur l'idée que la première a le pouvoir de produire la seconde. Or cette idée ne peut être fondée que sur une induction : il faut voir plusieurs fois l'effet se produire après l'apparition de la cause, pour conclure que celle-ci est à l'origine de l'effet. Jusqu'ici, l'argument de Hume est acceptable. Ce qui pose problème, c'est la conséquence qu'il en tire à propos de la liberté. Hume semble supposer que cette induction est impossible si l'effet n'apparaît pas à *chaque fois* que la cause apparaît. Faut-il que la conjonction de l'effet et de la cause soit observée à chaque fois ? Si elle ne se produisait que dans 95 % des cas, cela changerait-il quelque chose ? Ce serait sans doute suffisant pour croire qu'il y a une relation causale entre les deux faits. D'ailleurs, une cause toute seule n'est presque jamais suffisante. Elle ne peut opérer que dans certaines conditions ou avec le concours d'autres causes. Par exemple, on sait que lorsque la lumière du soleil augmente, la température augmente, toutes choses égales par ailleurs. Mais si un vent du nord se lève au moment où la lumière augmente, la température baisse. Aussi, l'observation d'une conjonction entre la lumière du soleil et son effet (la chaleur) n'est pas d'abord l'observation d'une conjonction *constante*, mais d'une conjonction *fréquente*. En physique et dans les sciences de la nature, l'observation amène les scientifiques à reformuler les lois pour qu'elles soient valables dans *tous les cas*, mais cet objectif est impossible ou très difficile à atteindre, et ils doivent souvent appliquer les lois avec précaution, en tenant compte des facteurs extérieurs qu'elles négligent, sans quoi les calculs sont faux ou inexacts.

¹¹ *Ibid.*, section VII, deuxième partie.

Par exemple, pour calculer la trajectoire d'un corps lancé en l'air, il faut appliquer les lois de Newton en tenant compte de la direction et de la force du vent, ainsi que du frottement de l'air qui limite l'accélération du corps dans sa chute. De plus, le frottement de l'air ne se manifeste pas de la même façon chez des corps de formes différentes : si un objet en forme d'hélice ne tombe pas tout droit, mais en tournant sur lui-même, c'est justement à cause du frottement de l'air, et il ne se comporterait pas de la même façon s'il n'y avait pas d'atmosphère. Les relations entre les causes et les effets sont considérées comme des relations causales même si elles ne sont pas observées dans tous les cas. L'argument épistémologique de Hume est donc fragile. Nous observons des relations plus ou moins régulières entre les causes et les effets, mais le fait qu'elles ne soient jamais d'une régularité parfaite ne nous empêche pas de supposer qu'il s'agit de relations causales.

En outre, il nous arrive souvent de postuler l'existence d'une certaine cause sans l'avoir observé *plusieurs fois* avec son effet, ou même sans l'avoir observé *une seule fois*. L'observation d'une conjonction constante ou régulière entre un effet et une cause n'est donc pas le seul point de départ pour découvrir une relation de causalité. Comme l'écrit Anscombe,

il n'est pas difficile de montrer qu'il est *prima facie* faux d'associer de cette manière la notion de cause à la nécessité ou à l'universalité. Car puisqu'il est bien plus facile de remonter des effets aux causes avec certitude que de prédire les effets à partir des causes, nous connaissons souvent une cause sans savoir s'il y a une loi générale de cette sorte, c'est-à-dire sans exception, ou s'il y a une nécessité. Par exemple, nous avons découvert qu'une certaine maladie est contagieuse. Si, dans cette situation, j'ai eu un contact (et un seul) avec une personne souffrant de cette maladie, et je deviens malade à mon tour, nous supposons qu'elle me l'a transmise. Mais que se passe-t-il si, ayant été en contact avec un malade, je demande au médecin si je vais devenir malade ? Normalement, il sera seulement capable de dire « Je ne sais pas : peut-être le serez vous, et peut-être ne le serez-vous pas ».¹²

Dans le cas de certaines actions libres, on peut admettre que l'effet est toujours précédé de la cause, mais cela ne veut pas dire que l'action est l'effet du déterminisme ou de la nécessité. Par exemple, l'existence d'un livre dont les caractères forment un texte intelligible et cohérent est impossible sans un auteur qui en est la cause. Mais il ne s'ensuit pas que l'agent devait *nécessairement* écrire ce livre. C'est souvent ainsi que nous découvrons les

¹²G. E. M. Anscombe, "Causality and Determination", leçon inaugurale à l'université de Cambridge, 1971, in *Metaphysics and the Philosophy of Mind*, vol. II de ses *Collected Philosophical Papers*, Basil Blackwell, Oxford, 1981. Ce passage se trouve p. 135-136.

causes de façon empirique : en constatant que tel effet ne survient jamais sans telle ou telle cause. Or cela n'exclut pas l'idée de *cause libre*.

Elizabeth Anscombe poursuit sa critique de Hume en rejetant le présupposé selon lequel nous ne pouvons pas découvrir d'emblée une relation de causalité dans un cas singulier :

La vraie – mais inutile – réponse à la question : « Comment obtenons-nous notre première connaissance de la causalité ? » est qu'en apprenant à parler, nous avons appris le sens et l'usage de tout un ensemble de concepts. Beaucoup d'entre eux se sont présentés sous la forme de verbes transitifs ou d'autres verbes d'actions employés pour décrire les faits observés. Les autres (comme on le voit bien avec le verbe « infecter ») ne forment pas des énoncés d'observation, mais expriment plutôt des hypothèses causales. Le mot « cause » en lui-même est très général. Comment un individu manifeste-t-il sa compréhension du concept de cause ? Nous sommes tentés de dire : par le simple fait que ce mot appartienne à son vocabulaire. S'il en est ainsi, alors la possession manifeste du concept présuppose la maîtrise de beaucoup d'autres mots. Ce que je veux dire ainsi, c'est que le mot « cause » ne peut être ajouté qu'à un langage dans lequel se trouvent déjà beaucoup de concepts causaux. Citons-en quelques-uns : *frotter, pousser, mouiller, transporter, manger, brûler, renverser, retenir, écraser, faire* (par exemple, des bruits, ou des bateaux en papier), *heurter*. Mais si nous essayons d'imaginer des langages dans lesquels il n'y a aucun concept causal particulier, alors aucune description de l'usage d'un mot dans ce langage ne peut le présenter comme signifiant « cause ». ¹³

En somme, on acquiert la maîtrise d'un concept abstrait et général comme celui de cause, grâce à la maîtrise préalable de concepts moins généraux comme brûler, détruire, fabriquer, lancer, pousser, etc. On acquiert la maîtrise de ces concepts causaux en voyant agir telle ou telle cause. Notre compréhension du mot « cause » est d'abord liée à l'observation de causes singulières, et non à la répétition d'une même expérience. Anscombe prend ainsi le contre-pied de la thèse de Hume. Il est facile, en effet, d'imaginer des situations où l'on observe directement une cause singulière produisant son effet. Quand on voit une personne allumer un feu, on voit la flamme de l'allumette se transmettre au papier, puis au tas de bois. Le foyer est plus grand à la fin, mais il naît en un mouvement continu que l'on peut observer étape par étape, de sorte qu'il est évident, même pour un observateur ignorant qui n'a jamais vu le feu, que la cause du feu de bois est bien la petite flamme de l'allumette. Dans un cas comme celui-ci, il n'est pas nécessaire de répéter l'expérience. Dès la première fois, on voit bien que la cause est la flamme de l'allumette. La même remarque vaut dans les situations ou

¹³*Ibid.*, p. 137.

un corps en pousse un autre, si le mouvement peut être observé en continu lorsqu'il se transmet à l'effet. Les verbes choisis par Anscombe à titre d'exemples signifient des actions directement observables. Quand un enfant apprend à parler, il apprend d'abord des mots désignant des choses ou des faits perceptibles. En grandissant, il acquiert progressivement des notions plus générales et donc plus éloignées des données sensibles, comme les concepts scientifiques ou philosophiques. Il est évident qu'il apprend les mots « brûler », « couper », ou « frotter », bien avant les mots « cause », « effet », ou « condition ». Mais comment acquiert-il ces concepts ? Tout simplement en écoutant d'autres personnes parler, car dans tous les cas où une chose pousse, frotte, coupe, brûle, ou renverse une autre chose, on dit qu'elle est la cause de ce qui se passe. Ainsi, parce qu'il entend d'autres personnes utiliser le mot « cause » dans un certain type de contextes, il acquiert insensiblement le concept de cause et devient capable de l'utiliser¹⁴.

Rien ne permet de dire que la connaissance des relations causales s'identifie à la connaissance de certaines relations nécessaires entre une catégorie de causes et une catégorie d'effets. Nous pouvons postuler ou observer une relation causale *singulière*. Il n'y a donc aucune raison d'ordre conceptuel pour supposer que toute relation causale est nomologique. Puisque cette affirmation n'a pas de fondement conceptuel, elle ne peut se justifier qu'au moyen d'une preuve empirique. Je vais soutenir maintenant que cette preuve n'existe pas, et qu'il n'y a donc aucun argument solide en faveur du déterminisme.

¹⁴ Il y a encore une manière de justifier l'idée que toute relation causale est nomologique : en affirmant que tout effet vient d'un ensemble de causes à la fois nécessaire et suffisant pour le produire. Appelons cette proposition le principe de la cause totale. D'après Hobbes, si toutes les causes et conditions nécessaires et suffisantes pour produire un effet sont réunies, alors il est impossible qu'elles ne le produisent pas. Si cet effet ne survient pas, c'est toujours parce qu'il manque une cause nécessaire à sa production. Cf. Thomas Hobbes, *De la liberté et de la nécessité*, p. 85 et s. Cette idée, malgré son évidence apparente, est douteuse. Anscombe la met en doute, mais sans donner une preuve de sa fausseté. Ted Honderich lui reproche cette attitude quelque peu méprisante (cf. Ted Honderich, *Mind and Brain*, Clarendon Press, Oxford, 1988, p. 62). Néanmoins, Honderich ne donne aucune preuve en faveur du principe. Or pour un libertarien, certaines causes - les personnes - ont souvent à la fois le pouvoir de produire un effet et celui de ne pas le produire. Selon cette idée, le même ensemble de causes suffisantes peut produire son effet, ou bien ne pas le produire. Pour un libertarien, il est donc évident que si un déterministe s'appuie sur le principe de la cause totale pour justifier la thèse déterministe, il commet une pétition de principe. Si nous avons des intuitions fortes qui s'opposent à ce principe (les intuitions sur la responsabilité, l'existence de futurs contingents, etc.), la charge de la preuve incombe aux déterministes, car il ne bénéficie pas d'une évidence aussi grande *a priori*. Il paraît solide quand on l'applique aux êtres inertes ou qui n'ont pas le libre arbitre, mais rien ne nous oblige à l'appliquer aux agents humains. Puisqu'il n'est pas fondé sur une évidence conceptuelle, il faut y renoncer, ou le prouver empiriquement.

7.4 Le déterminisme peut-il avoir un fondement empirique ?

J'ai montré, avec Anscombe, que les relations de cause à effet ne se présentent pas nécessairement sous la forme de relations parfaitement régulières, ni même générales. Le concept de cause n'implique pas que toute relation causale soit nécessaire. Ainsi, la prémisse (3) de l'argument déterministe n'est ni immédiatement évidente, ni fondée sur une preuve conceptuelle. La seule manière de la justifier serait de la prouver par induction en observant les phénomènes naturels. Peut-on prouver par induction que toute relation causale soit nécessaire ? Pour montrer que c'est a priori impossible, j'adopterai une stratégie qui est l'inverse de la précédente : je soutiendrai qu'une relation *constante* observée dans la nature n'est pas nécessairement une *loi* de la nature ni une relation *causale*. Comme l'écrit Swinburne,

Tous les corbeaux observés jusqu'à maintenant sont noirs. Mais ceci n'est en rien une preuve adéquate pour supposer une loi de la nature selon laquelle les corbeaux sont noirs. Nous avons besoin d'une preuve pour supposer que si tous les corbeaux observés jusqu'à maintenant sont noirs, ce n'est pas à cause d'un simple accident lié aux conditions locales (c'est-à-dire dû à une certaine chance qu'il n'y ait pas de mutation produisant un gène de la blancheur dans le génotype du corbeau) [...]

Si les hommes du dix-huitième siècle ont cru que les lois newtoniennes de la gravité (selon lesquelles tous les corps matériels s'attirent mutuellement avec des forces proportionnelles au produit de leurs masses et inversement proportionnelles au carré de la distance qui les sépare) étaient effectivement des lois de la nature, c'est parce qu'elles étaient des lois très simples s'intégrant dans une théorie à quatre lois en tout, les autres étant également simples, et parce qu'on a vu qu'ensemble, elles étaient capables de prédire une grande variété de phénomènes.¹⁵

Une simple généralisation ne permet pas de découvrir une loi de la nature. Le fait que deux propriétés semblent aller toujours de pair peut être dû à des conditions particulières dans lesquelles toutes les observations ont été faites. Par exemple, les propriétés génétiques des corbeaux ne sont peut-être pas la cause suffisante de la couleur noire observée dans toutes les populations de corbeaux. Pour l'affirmer, il faut d'abord tester cette hypothèse explicative. Si elle résiste à tous les tests expérimentaux, elle sera considérée comme valable, parce qu'elle aura été corroborée par l'expérience dans des conditions suffisamment variées. En effet, le but de l'expérimentation est de la tester dans des conditions aussi variées que possible. Ensuite,

¹⁵ Richard Swinburne, *The Evolution of the Soul*, Oxford University Press, Oxford, 1997 (édition revue et corrigée), p. 187.

l'utilisation fructueuse d'une hypothèse pour la prédiction de phénomènes est encore un test, même si son but premier n'est pas de tester une hypothèse. Plus l'hypothèse permet de faire des prédictions, plus elle a de chances d'être vraie. La crédibilité de l'hypothèse dépend de son pouvoir explicatif et prédictif.

Ainsi, il est clair qu'une induction ne se fonde pas seulement sur la répétition d'une observation. Cette répétition n'autorise pas toujours une généralisation. Par exemple, si en fouillant dans ma poche, j'y trouve une pièce de cinq centimes, puis une deuxième, une troisième, etc., cela ne prouve pas que tous les objets qui sont dans ma poche soient des pièces de cinq centimes. La généralisation se fonde à la fois sur l'observation d'une régularité et sur une explication de cette régularité, car c'est en connaissant sa cause que nous pouvons savoir jusqu'où elle s'étend. Reprenons l'exemple de Swinburne. Les lois de Kepler décrivent des régularités observées dans les trajectoires des planètes et des comètes¹⁶. Newton a formulé une explication de ces régularités en énonçant de nouvelles lois du mouvement des corps. À partir de ces lois, on peut inférer que *partout dans l'univers*, les corps en orbite doivent se mouvoir de la même façon que les planètes du système solaire. L'explication proposée par Newton en énonçant les lois du mouvement des corps permet donc de généraliser l'observation faite sur les planètes du système solaire à tous les corps en orbite (en tenant compte des éventuelles perturbations liées à la présence d'autres corps). Je n'ai pas l'intention d'énoncer tous les critères de la confirmation empirique, mais cela n'est pas nécessaire pour mon propos. Il est clair que la confirmation inductive d'une affirmation générale dépend d'autre chose que la simple répétition d'une observation. Les lois de Newton ont été admises comme vraies parce qu'elles fournissaient une explication simple et cohérente de nombreux phénomènes observables, et parce qu'elles permettaient de prédire les mouvements des corps.

¹⁶ Kepler décrit le mouvement des planètes par les trois lois suivantes : 1) L'orbite de chaque planète a la forme d'une ellipse dont le Soleil occupe l'un des foyers ; 2) la droite joignant la planète au Soleil balaie des aires égales en des temps égaux ; 3) le carré du temps de révolution autour du Soleil de n'importe quelle planète (y compris la Terre) est proportionnel au cube de sa distance moyenne au Soleil. Cf. I. Bernard Cohen, *Les origines de la physique moderne*, trad. fr. J. Métadier et G. Jeanmougin, éditions du Seuil, 1985, Paris, p.155-160 (édition originale : *The Birth of a New Physics*, W. W. Norton & Company, 1960, New York). La première loi de Kepler reçoit son explication physique dans les propositions I, II et XI, des *Principia* de Newton : « Newton démontre dans le premier théorème du livre I des *Principia* que, si un corps est constamment attiré vers un centre de force, son mouvement, qui sans cette influence serait inertiel, devient un mouvement curviligne, et toute ligne droite joignant le corps au centre balaie des aires égales en des temps égaux. Dans la proposition II (Théorème II), il démontre que, si un corps se déplace le long d'une courbe de sorte que les aires balayées par une droite joignant le corps à un point quelconque sont proportionnelles aux temps de balayage, il existe une force « centrale » (centripète) attirant continuellement le corps vers ce point. La signification de la première loi de Kepler n'apparaît qu'à la proposition XI, dans laquelle Newton cherche « la loi de la force centripète lorsqu'elle tend à un foyer de l'ellipse ». Cette force varie « inversement comme le carré de la distance ». » (*Ibid.*, p.182-183)

Peut-il exister une preuve du même genre en faveur de la thèse déterministe ? Plus précisément, y a-t-il une preuve inductive en faveur de l'idée que toute relation de cause à effet est nomologique ? C'est peu probable.

Premièrement, cette hypothèse n'augmente nullement notre capacité de prédire des phénomènes. En effet, c'est une affirmation générale au sujet des relations de causalité. C'est donc une loi au sujet des lois naturelles, un loi de second ordre, en quelque sorte. En admettant qu'elle soit vraie, je pourrais la connaître sans avoir les moyens de prédire aucun phénomène physique, car ce sont les lois de premier ordre, c'est-à-dire les lois physiques, qui permettent de prédire les phénomènes physiques. Si j'ignore ces lois, la connaissance d'une *méta-loi* portant sur l'ensemble des relations de cause à effet ne compensera pas ce manque. Inversement, la connaissance des lois de premier niveau (lois physiques, biologiques, ou autres) est la seule utile pour prédire les phénomènes, et on ne voit pas pourquoi la connaissance d'une loi plus abstraite au sujet des relations de causalité faciliterait cette activité de prédiction.

Deuxièmement, une hypothèse aussi abstraite ne peut servir à expliquer les phénomènes, car au lieu de montrer la cause d'un effet observable, elle affirme quelque chose au sujet de toutes les relations de cause à effet. En admettant que cette affirmation générale soit vraie, elle ne montre la cause d'aucun effet. C'est pourquoi elle n'en explique aucun.

Les sciences de la nature expliquent des régularités à partir d'autres régularités, comme nous le voyons dans l'exemple de Kepler et Newton : les lois de Newton apportent l'explication des lois de Kepler. Mais cela ne prouve pas qu'il faille s'élever jusqu'à une loi universelle ou jusqu'à une loi de second ordre. Bien que la régularité des phénomènes soit importante du point de vue de l'explication scientifique, elle ne requiert pas que tout événement physique soit déterminé causalement. Si un événement x d'une probabilité faible ou moyenne cause un événement y , l'explication de y à partir de x est peu éclairante. Le caractère *régulier* du phénomène auquel nous nous référons pour expliquer un autre phénomène est donc très important du point de vue explicatif. Mais l'explication des phénomènes naturels ne suppose pas que toutes les relations de cause à effet soient *nécessaires*.

Ainsi, l'induction peut nous faire découvrir des vérités générales sur la cause de certains effets, mais rien ne permet de dire qu'on puisse faire un pas de plus dans l'abstraction et dans

la généralité en supposant que toutes les lois sont nécessaires, ou que toutes les relations de cause à effet sont nécessaires. Cette induction (ou cette « méta-induction ») ne peut être soumise à des tests empiriques, et la régularité des relations causales observées dans les phénomènes ne peut être généralisée à *toutes* les relations causales. Pour cela, il faudrait que cette généralisation modifie notre capacité de prédire les phénomènes, or il n'y a aucune raison pour qu'elle la modifie.

Cependant, il est parfois légitime de faire une hypothèse qui n'augmente pas notre pouvoir de prédiction, si cette hypothèse fournit l'explication la plus simple des phénomènes observés. Il faut donc voir si le déterminisme, ou l'idée selon laquelle toute relation causale est nécessaire, peut se fonder sur une inférence à la meilleure explication. Cette inférence s'appuie sur un critère de vraisemblance appliqué aux causes d'un phénomène : si on explique le phénomène à partir d'un grand nombre de causes ou de facteurs, la conséquence évidente de l'hypothèse est qu'il était extrêmement improbable car il supposait le concours d'une multitude de causes indépendantes. L'hypothèse explicative la meilleure est celle qui postule le plus petit nombre de causes (à condition qu'il s'agisse de causes ordinaires, et non de lutins, de fées, etc.). Par exemple, lorsqu'on voit des traces de pas dans la neige, disposées de telle sorte qu'un individu a pu les faire tout seul en marchant normalement, on n'explique pas ces traces en imaginant un individu marchant à reculons ou à cloche-pied. Ces deux hypothèses supposent le même nombre de causes que la première, mais renvoient sans nécessité à des situations peu ordinaires, donc moins vraisemblables que celle où un individu marche normalement. On peut imaginer aussi quatre individus sautant à cloche pied, ou mettant soigneusement leurs pas dans les empreintes laissées par le premier d'entre eux, mais c'est un processus qui requiert plus de causes et qui paraît donc plus difficile à obtenir. Là aussi, on renvoie sans nécessité à des situations improbables.

Le déterminisme pourrait être défendu en tant qu'hypothèse la plus probable, pour ces deux raisons à la fois : il suppose moins de causes ou de catégories d'entités différentes ; ensuite, il renvoie à des phénomènes familiers, bien connus des sciences de la nature. Certains libertariens supposent que la cause ultime des actions dont nous sommes responsables, c'est nous-mêmes en tant qu'agents. Cette cause n'est donc pas un événement, mais l'agent lui-même, qui est plutôt une substance. En fait, comme nous le verrons plus loin, il y a deux théories différentes chez les libertariens ; certains, comme Kane, pensent que la cause ultime des actions libres est un ensemble de processus quantiques indéterminés. Il s'agit donc

d'événements, c'est pourquoi on l'appelle théorie de la "causalité événementielle", par opposition à la première, dite "théorie de la causalité de l'agent". La théorie de Kane pose des difficultés sur lesquels il nous faudra revenir. En admettant que la meilleure théorie libertarienne soit l'autre, elle suppose deux catégories de relations causales : d'une part, les relations entre événements, bien connues des sciences de la nature et dont on admettra facilement l'existence, et d'autre part, ce que j'appellerai les relations agent-événement, où la cause d'un événement est un agent, compris comme irréductible à un ensemble d'événements. Cette notion pose un problème : est-il possible de concevoir une relation causale où l'agent soit cause, et qui ne soit pas en même temps une relation entre des événements, par exemple la décision de l'agent et ses mouvements corporels ? Que cette relation existe ou non, elle paraît en tout cas moins familière que l'autre, dont l'existence ne fait aucun doute. Aussi, en affirmant l'existence des relations causes-événements, on choisit une hypothèse apparemment moins simple que celle des déterministes. Ceux-ci pourraient en conclure qu'elle est moins probable, mais ce serait aller un peu vite en besogne, car la simplicité à évaluer est aussi celle des révisions impliquées par chaque hypothèse dans l'ensemble de nos croyances ou de nos intuitions, et sur ce point, l'avantage semble appartenir plutôt aux libertariens. Il y a une grande différence entre une hypothèse qui s'accorde avec nos intuitions sur la responsabilité morale, et une autre qui conduit à les rejeter ou à les réviser profondément. Au contraire, l'idée des relations agents-événements n'est pas improbable au sens où peut l'être une hypothèse trop complexe, car elle ne suppose pas le concours d'un grand nombre de causes, de facteurs ou de conditions. Bien compris, le principe de simplicité ne fournit pas un argument en faveur du déterminisme, mais plutôt en faveur du libertarisme. La thèse déterministe aboutit à une dissociation de l'être et de l'apparence. En effet, nous nous attribuons spontanément la responsabilité morale, et un pouvoir d'agir autrement qui implique de la contingence au sens fort du terme. Si le déterminisme exclut cela, comme je l'ai soutenu dans les chapitres précédents, il aboutit à l'idée que nous ne sommes pas ce que nous croyons être. C'est donc une théorie qui manque de simplicité au sens où elle demande une révision des croyances communes sur la liberté et la nature humaine.

Ainsi, aucune preuve empirique ne justifie l'idée que toute relation de cause à effet est une relation nécessaire, et puisque cette idée est un des piliers de la thèse déterministe, le déterminisme n'est pas confirmé par l'observation de la nature ; c'est une croyance fondée sur des principes métaphysiques, ou sur une image du monde physique suggérée par la culture dans laquelle nous vivons. Or dans les sections 7.2 et 7.3, j'ai soutenu que le déterminisme n'a

pas de fondement *a priori* ou conceptuel. Je ne vois donc aucune raison d'accepter cette affirmation si générale et si catégorique. Si elle n'a ni fondement empirique, ni fondement conceptuel, c'est qu'elle n'a pas de fondement.

Une position plus modeste serait de dire que notre connaissance du cerveau humain exclut l'existence de relations causales non-nécessaires à l'intérieur du cerveau. Mais cette thèse défendue par Ted Honderich¹⁷ ne s'impose nullement, car les scientifiques eux-mêmes admettent souvent l'existence de phénomènes quantiques à l'intérieur du cerveau¹⁸. Un philosophe peut difficilement s'appuyer sur les données scientifiques pour argumenter au sujet du libre arbitre. Ce serait peut-être envisageable s'il y avait un consensus assez grand, parmi les scientifiques, sur les processus qui interviennent dans la formation des intentions et des actions volontaires. Mais puisqu'un tel consensus n'existe pas, les philosophes doivent se contenter d'arguments métaphysiques ou épistémologiques d'une portée plus générale.

¹⁷ Ted Honderich, *A Theory of Determinism*. Dans le premier volume (*Mind and Brain*), il passe en revue les principales théories sur le rapport entre les états mentaux et les états physiques du cerveau, puis défend sa propre théorie matérialiste en l'appuyant sur la connaissance scientifique du cerveau humain. Selon lui, les neurosciences mettent en évidence deux faits importants du point de vue philosophique : la proximité entre les états mentaux et neuronaux (« *psychoneural intimacy* »), et l'existence de relations causales nomologiques entre chaque activité de l'esprit et un certain événement dans le cerveau. La première idée est difficilement contestable, car les moyens dont nous disposons maintenant pour observer le cerveau en activité nous donnent des indications assez précises sur les phénomènes qui se produisent dans le cerveau lors des différentes activités (calcul, lecture, audition, imagination, effort de mémorisation, etc.). Le point à discuter est donc le deuxième, ou plutôt la conclusion qu'en tire Honderich. En effet, un libertarien peut reconnaître qu'il y a des relations causales nomologiques entre les états mentaux et les états physiques du cerveau, sans admettre l'idée que *toutes* les relations causales entre l'esprit et le cerveau sont nomologiques. Selon Honderich, l'observation nous fait connaître ceci :

(i) Dans de très nombreux cas, un ensemble *n* de faits neuronaux est souvent, ou très souvent, ou dans toutes les expériences et les observations, accompagné d'un certain phénomène mental *m*. [...]

(ii) Dans de très nombreux cas, un ensemble *n* de faits neuronaux apparaît comme provoquant à coup sûr un phénomène mental *m*. (p. 284)

Pour Honderich, ces observations prouvent que la causalité dans le cerveau est parfaitement régulière et nomologique. Mais cette conclusion est discutable pour plusieurs raisons. Premièrement, le cerveau humain est d'une complexité immense: d'après les estimations récentes, il se compose d'environ cent milliards de neurones ; et selon Honderich lui-même, la plupart des neurones « ont de très nombreuses connections synaptiques, entre 1000 et 10000, ou plus » (p. 263). Il est donc très difficile d'avoir une connaissance à peu près complète des relations causales existant à l'intérieur du cerveau. Deuxièmement, certains scientifiques affirment qu'il y a dans le cerveau humain des relations causales non-nécessaire, mais probabilistes, à savoir des processus quantiques. cf. note suivante.

¹⁸ En 1984, Margenau a formulé une théorie sur les phénomènes quantiques dans le cerveau, qui a inspiré Beck et Eccles pendant les années 90. Quand Honderich critique Popper et Eccles, il se réfère à la première théorie de Eccles, celle de 1977. La théorie proposée par Eccles dans ses dernières publications repose sur des données empiriques bien différentes de celles invoquées par Honderich. En effet, Eccles déclare que la transmission des influx nerveux au niveau synaptique (par l'émission de molécules de transmission) est d'une probabilité variable, et parfois très faible : « il est possible que certains boutons ne soient pas éloignés d'une probabilité de 1, mais on n'a encore jamais observé de cas où la probabilité était égale à l'unité ».

L'argument que j'ai discuté est de cet ordre, mais ses prémisses sont douteuses. Rappelons la forme de l'argument :

- (a) Toute action est un événement.
- (b) Tout événement a une cause.
- (c) Toute relation de cause à effet est une relation nécessaire.
- (d) Donc toute action est causée de façon nécessaire.

En 7.2, après avoir discuté plusieurs formulations du principe de causalité, j'ai retenu la formulation employée ici dans la deuxième prémisses. Les prémisses (1) et (2) semblent correctes, mais la troisième est douteuse, car elle n'a ni fondement conceptuel, ni fondement empirique. Contrairement à ce que déclare Hume¹⁹, le concept de cause n'est pas toujours lié à l'observation de « connexions constantes » entre deux phénomènes. La prémisses (3) n'a donc pas de fondement *conceptuel*. Or on ne peut connaître de manière *empirique* que toutes les relations de cause à effet sont nécessaires, car si l'induction permet de découvrir certaines relations causales, elle ne permet pas d'énoncer des lois au sujet de *toutes* les relations causales. Pour que ce pas soit justifié, il faudrait que l'hypothèse selon laquelle toutes les relations causales sont nécessaires augmente notre pouvoir de prédiction et notre connaissance des causes. Je pense donc que la thèse déterministe, non seulement ne s'impose pas, mais est difficilement soutenable. Par conséquent, il me semble que ce qui menace la thèse libertarienne est beaucoup moins la force des arguments déterministes que le côté mystérieux du libre arbitre. Dans la suite de ce chapitre, je montrerai pourquoi cette faculté est mystérieuse. Après avoir introduit brièvement la notion d'indéterminisme physique invoquée dans les débats contemporains, j'aborderai plusieurs difficultés liées à cette notion. La thèse libertarienne suppose que les actions libres ont leur origine dans des choix ou des événements indéterminés. Or il n'est pas facile de voir comment une action peut être rationnelle sans être déterminée par des raisons. Ensuite, on se demande comment les événements physiques situés à la source des actions peuvent être contingents tout en étant contrôlés par l'agent.

¹⁹ Cf. *supra*, p. 332.

7.5 Le hasard et l'indéterminisme

Dans la tradition philosophique, on peut distinguer deux conceptions dominantes du hasard qui s'opposent entre elles. Pour les philosophes atomistes comme Démocrite et Épicure, le hasard (« *automaton* ») est le mouvement qui est à l'origine de la formation du monde. Les atomes tombent dans le vide de toute éternité selon un mouvement rectiligne, puis la déviation spontanée (c'est-à-dire sans cause extérieure) d'un atome, déclenche un mécanisme aboutissant à la formation du monde. Dans cette perspective, le hasard, comme le suggère le mot grec « *automaton* », se caractérise par l'absence de cause ou de déterminisme causal.

Selon l'autre conception du hasard, un événement qui se produit par hasard n'est pas sans causes, mais n'a été ni prévu ni souhaité. Quand cet événement est heureux, on parle de « fortune » car il ressemble à un événement provoqué volontairement, comme si les dieux l'avaient voulu. Par exemple, si un homme qui va au marché y rencontre accidentellement un de ses débiteurs en train de recevoir de l'argent, c'est un heureux hasard, mais l'événement a une cause. Dans cet exemple, le hasard est la rencontre imprévue de plusieurs chaînes causales : celle qui pousse le débiteur à aller au marché, et celle qui pousse le prêteur à s'y rendre. Cette deuxième conception du hasard est compatible avec l'idée que rien n'est sans cause ou que tout est soumis à des lois nécessaires. La première, au contraire, suppose l'existence de mouvements incausés ou indéterminés, et c'est justement cette idée qui, dans l'épicurisme, permet d'accorder une place au libre arbitre. Chez certains philosophes contemporains, comme Kane, le recours à la notion de processus physique indéterminé remplit à peu près la même fonction.

En physique quantique, la notion d'indéterminisme a été interprétée de deux façons. La controverse entre les tenants des interprétations objectiviste et épistémique a opposé Bohr et Einstein²⁰, et s'est prolongée bien après eux. Le point de départ du problème est l'expérience du microscope à rayons gamma imaginée par Heisenberg et la formulation des « relations

²⁰ Cf. le fameux article dit « EPR » en référence aux initiales de ses auteurs : Albert Einstein, Boris Podolsky, et Nathan Rosen, « Can Quantum-Mechanical Description of Physical Reality Be Considered Complete ? », *Physical Review* 47 (1935), p. 777-780. Niels Bohr y répondit dans le numéro suivant de la *Physical Review*, p. 696 et s.

d'incertitude »²¹. Heisenberg a montré qu'il est impossible de mesurer à la fois la vitesse et la position d'une particule. Soit on l'éclaire, et on perturbe sa trajectoire, soit on ne l'éclaire pas et on ne peut rien savoir de sa trajectoire. L'éclairage soumet la particule à une pression de radiation qui la dévie en la propulsant selon la direction de la lumière. Il est en fait possible de mesurer successivement la vitesse et la position d'une particule, mais le résultat de ces mesures varie selon l'ordre dans lequel on procède : en commençant par la mesure de vitesse, on n'obtient pas le même résultat qu'en commençant par la mesure de position. D'une manière plus générale, le problème vient du fait que la matière se comporte tantôt comme si elle était composée de corpuscules, tantôt comme si elle était composée d'ondes. En conséquence, la physique quantique attribue aux particules de matière des propriétés ondulatoires (fréquence et longueur d'onde) et des propriétés corpusculaires (énergie et impulsion). Or, un objet qui a une fréquence et un vecteur d'onde ne peut avoir de localisation précise. Une onde sonore, par exemple, n'a pas de localisation très précise. De même, on ne peut connaître la fréquence d'une onde à un instant précis, car la détermination de la fréquence suppose d'échantillonner le signal pendant une certaine durée.

Ces données ont fait l'objet du conflit entre l'interprétation objectiviste, selon laquelle les particules sont *réellement* indéterminées, et l'interprétation épistémique, selon laquelle l'indétermination relève seulement d'une incertitude liée aux limites de nos instruments d'observation²². Actuellement, l'interprétation objectiviste est dominante. D'après Étienne Klein, « la bonne façon d'interpréter le principe de Heisenberg consiste à dire, non pas qu'il est impossible de déterminer simultanément la position et l'impulsion des particules, mais bien plutôt à affirmer que ces dernières ne possèdent jamais ces deux attributs simultanément. Plus précisément, la représentation formelle que la physique quantique se fait des particules ne leur attribue jamais ces deux caractéristiques à la fois »²³. Mais si c'est le formalisme de la

²¹ Cf. par exemple, Louis de Broglie, *La Physique nouvelle et les quanta*, Flammarion, Paris, 1937, réimp. dans la collection Champs-Flammarion, p. 213-229.

²² L'interprétation objectiviste est sévèrement critiquée par Mario Bunge dans *Philosophy of Physics*, D. Reidel Publishing Company, 1973, Dordrecht, Hollande (trad. fr. Françoise Balibar : *Philosophie de la physique*, Seuil, Paris, 1975). Bunge lui reproche de se fonder sur une philosophie périmée : le positivisme logique de l'entre-deux-guerres. Cf. ch. 5 (p. 114-134 dans l'édition française).

²³ É. Klein, *Petit voyage dans le monde des quanta*, Champs-Flammarion, Paris, 2004. cf. aussi Louis de Broglie, *La physique nouvelle et les quanta*, Flammarion, Paris, 1937, réédité en 1973. Voir le ch. X., en particulier p. 213-217. Louis de Broglie explique de deux manières différentes la cause de l'indétermination dans les mesures. Il n'adhère pas à l'interprétation de l'école de Copenhague (qui deviendra dominante), selon laquelle l'indétermination doit être considérée comme réelle.

physique quantique qui implique que les particules sont indéterminées, on se demande s'il décrit fidèlement la réalité ou s'il faut le voir comme un appareil conceptuel à la fois commode et partiellement faux. De nos jours, certains²⁴ voient encore dans la notion d'indéterminisme la preuve que la physique quantique est une théorie incomplète, et de temps en temps, des physiciens s'efforcent de la remplacer par une nouvelle théorie purement déterministe. Devant ces difficultés, nous devons adopter la plus grande prudence face aux utilisations philosophiques de la notion d'indéterminisme.

7.6 Le libre arbitre et l'indéterminisme quantique. Première difficulté

Parmi les philosophes incompatibilistes, c'est Kane qui est le plus connu pour son recours à la notion de phénomène physique indéterminé. Son exposé le plus complet sur la place de l'indéterminisme quantique dans l'action libre se trouve dans *The Significance of Free Will* (1998). En réalité, Kane n'est pas le premier à utiliser cette idée. Dès les débuts de l'interprétation objectiviste, des scientifiques et des philosophes se sont interrogés sur ses enjeux métaphysiques. Par exemple, Arthur Eddington, dès 1928, a soutenu que l'indéterminisme indiquait une nouvelle voie pour comprendre la place du libre arbitre dans le monde physique²⁵. Mais bien vite, des objections se sont présentées. La plus importante est le *Mind argument*, ainsi nommé parce qu'il a été formulé plusieurs fois dans la revue *Mind*. Avant d'aborder cet argument philosophique, j'évoquerai brièvement ici une objection d'ordre scientifique.

Cette objection est que l'existence de phénomènes physiques indéterminés comparables à ceux décrits par la science contemporaine ne permet pas d'expliquer l'existence de processus indéterminés à un niveau macroscopique, car normalement, l'indéterminisme physique ne concerne que le niveau microscopique ou subatomique. Le fait que des particules traversent des barrières d'énergie de façon inexplicable et sans y être déterminées par aucune cause ne peut avoir qu'un effet négligeable au niveau macroscopique. Or c'est au niveau macroscopique que nous situons la liberté et les possibilités alternatives. La différence entre

²⁴ Ted Honderich est au moins tenté par cette idée. Cf. Ted Honderich, *Mind and Brain*, p. 320-325.

²⁵ Cf. Arthur Eddington, *The Nature of the Physical World*, McMillan, London, 1928.

deux possibilités comme prendre le train ou partir en voiture, est une différence macroscopique. Si les possibilités alternatives n'existaient qu'au niveau microscopique ou subatomique, elles n'auraient guère de pertinence du point de vue de la liberté, car elles ne nous empêcheraient pas d'être déterminés dans tous nos mouvements.

En réalité, cet argument n'est pas décisif, car la théorie du chaos a permis de montrer que dans de nombreux phénomènes physiques (les phénomènes « chaotiques »), une très légère perturbation entraîne rapidement des conséquences démesurées, apparemment disproportionnées si on les compare à l'évènement déclencheur. La théorie du chaos a donc une application intéressante dans le débat sur le libre arbitre. Cette application a été mise en évidence à la fois par des philosophes et par des scientifiques : Henry Stapp (1990), George P. Scott (1991), James Garson (1993), James Kellert (1993) ou encore Robert Kane. Comme le dit Garson,

les systèmes chaotiques sont, par définition, des amplificateurs de perturbations. Ils peuvent échapper à la règle selon laquelle les indéterminations quantiques se neutralisent au niveau macroscopique. Puisque les événements chaotiques dépendent exclusivement de très petites choses, le comportement des particules subatomiques dans une très petite région de l'espace-temps peut être répercuté jusqu'à donner des différences massives au niveau du comportement global. Si nous voyons le cerveau comme un système chaotique qui amplifie les événements subatomiques, alors l'imprédictibilité chaotique n'est plus seulement épistémologique ; elle est métaphysique (ou plutôt physique)²⁶.

Pour Stapp, les données les plus récentes suggèrent en effet que le cerveau est un système chaotique :

Les simulations classiques avec des ordinateurs montrent que l'évolution de leur état macroscopique est très sensible aux petites variations au niveau synaptique.²⁷

Au XIX^{ème} siècle déjà, Maxwell avait remarqué l'importance des phénomènes chaotiques dans la nature. Les équations déterministes ne pouvant être appliquées sans une connaissance précise de l'état initial du système, connaissance impossible en l'occurrence, la

²⁶ James Garson, « Chaos and Free Will », p. 4, conférence donnée à l'American Philosophical Association, Pacific Division meeting, march 1993, et citée par Kane dans *The Significance of Free Will*, p. 129.

²⁷ H. Stapp, « A Quantum Theory of the Mind-Body Interface », p. 10, conférence non publiée, Université de Californie à San Francisco, citée par Kane dans *The Significance of Free Will*, p. 129. John Eccles cite et commente la même conférence dans *Comment la Conscience contrôle le cerveau*, p. 73, ainsi que deux autres conférences de Henry Stapp.

thèse déterministe ne peut être confirmée à propos des systèmes chaotiques. Or si le déterminisme est une simple conjecture, il ne peut fonder aucune objection contre l'idée de libre arbitre.

Maxwell, évidemment, ne s'appuie pas sur la physique quantique, qui n'existe pas encore à son époque, mais sur l'exemple de systèmes dits « chaotiques » dont le comportement est imprévisible²⁸. Kane, de son côté, combine les deux références : il pense que la liberté suppose l'existence de phénomènes physiques indéterminés *à l'intérieur d'un système chaotique*. Bien qu'ils soient imprévisibles, les phénomènes chaotiques peuvent exister même là où vraisemblablement, tout est soumis à une nécessité implacable (par exemple, dans les phénomènes météorologiques). Une machine pourrait être assez complexe pour que ses mouvements soient partiellement imprévisibles. Elle constitue alors un système à la fois chaotique, et entièrement mécanique, donc soumis à des lois de type déterministe. Mais bien entendu, Kane ne pense pas que l'aspect chaotique d'un processus dans le cerveau soit suffisant pour qu'il y ait libre arbitre. Néanmoins d'après lui, c'est l'existence de phénomènes indéterminés dans le cerveau qui rend le comportement humain plus ou moins imprévisible, car si le cerveau est un système physique de type chaotique, il peut amplifier les indéterminations. Ainsi, une indétermination au niveau *microscopique* peut entraîner une indétermination au niveau *macroscopique*. D'où le caractère imprévisible du comportement humain. Par exemple, si contrairement à ses habitudes les mieux enracinées, un de mes amis renonce à prendre un café pendant la pause de 10h, cela pourrait être dû à un événement indéterminé au niveau subatomique.

²⁸ Cf. John Barrow, *La grande théorie*, trad. Michel Cassé, Loïc Cohen et Guy Paulus, Albin Michel, Paris, 1994 (puis Flammarion, Paris, 1996. Barrow évoque les idées de Maxwell p. 61-62). L'édition originale aux presses universitaires d'Oxford (1991) s'intitule *Theories of Everything. The Quest for Ultimate Explanation*.

7.7 Le *Mind argument* : la notion de hasard

Si l'idée de phénomènes indéterminés à l'intérieur du cerveau humain s'accorde bien avec le fait que nous agissons de manière contingente et imprévisible, elle pose plutôt un problème lorsqu'il s'agit de comprendre comment l'agent contrôle ses actions. L'argument du *Mind* consiste à dire que l'indéterminisme, au lieu de s'accorder avec le libre arbitre et la responsabilité morale de l'agent, s'y oppose :

Si les lois de la nature sont indéterministes, alors il y a en effet plusieurs futurs compatibles avec ces lois et avec ce qui est actuellement passé et présent, mais comment aurait-on la moindre possibilité de choisir que l'un de ces futurs devienne actuel ? N'est-ce pas seulement par l'effet du hasard qu'ils deviennent actuels ? (P. Van Inwagen, « The Mystery of metaphysical Freedom »²⁹)

Cet argument a été présenté de différentes manières. J'essaierai d'en examiner toutes les versions possibles, en m'inspirant surtout de Van Inwagen³⁰. Pour les incompatibilistes, il est évident que l'existence de processus indéterministes n'est pas une condition suffisante du libre arbitre, mais une condition nécessaire. Imaginons un monde entièrement soumis au déterminisme, à l'exception d'une particule qui évolue à des années-lumières de tout agent rationnel, de telle sorte que personne ne peut influencer sa trajectoire ou son comportement. Dans un tel monde, le libre arbitre n'existe pas malgré l'existence d'un phénomène indéterministe. Si nous imaginons un monde où de très nombreuses particules se déplacent de manière indéterministe, loin mais aussi tout près de nous, cela ne suffit pas non plus pour que nous ayons le libre arbitre. Pour que des phénomènes indéterministes permettent le libre arbitre, il faut qu'ils influencent nos actes ou les constituent.

Prenons donc un exemple d'action indéterminée, et voyons si l'indéterminisme peut favoriser le libre arbitre et la responsabilité morale de l'agent. Un voleur entre dans une petite église de campagne³¹. Il y voit un tronc servant à collecter de l'argent pour les pauvres. En l'ouvrant, il ne trouve qu'une somme misérable. Mais les affaires sont les affaires. Il décide donc d'empocher l'argent. Or soudain, une image lui revient à l'esprit : le visage suppliant de sa mère allongée sur son lit de mort, au moment où elle lui demande d'être toujours un homme

²⁹ p. 190 dans l'anthologie de Kane (2002a).

³⁰ Cf. Peter Van Inwagen, *An Essay on free will* (1983), p.126-152, et deux articles : « The Mystery of metaphysical Freedom » (1998), et « Free Will remains a Mystery » (2002).

³¹ Je reprends ici l'exemple de Van Inwagen dans *An Essay on Free Will*, p. 127-128.

honnête. Il se souvient alors de sa promesse, et renonce à voler l'argent. Ce n'est pas la première fois qu'il s'en souvient, mais c'est la première fois qu'il y prête attention. Après quelques instants de réflexion, il quitte l'église sans rien emporter.

Pour les incompatibilistes, ce voleur est responsable de son choix seulement si celui-ci est indéterminé. Selon cette idée, il y a des mondes possibles accessibles au voleur où tout se passe exactement comme dans le monde actuel jusqu'au moment du choix, et où le voleur fait néanmoins un autre choix. Une première version de l'argument du *Mind* consiste à dire qu'une action indéterminée comme celle du voleur est un événement qui relève du *hasard* ou de la *chance*, et qu'un tel événement ne peut être une action libre.

Van Inwagen répond à cette objection en observant que le mot « hasard » (« *random* ») s'applique normalement à des *suites d'événements*, comme dans les jeux de hasard ou dans une table de nombres aléatoires. Le sens de l'objection pourrait être que si la thèse libertarienne est vraie, le fait de décider quelque chose librement est comparable au fait de prendre une décision après avoir consulté une table de nombres aléatoires. Mais la notion d'indéterminisme n'est pas toujours comprise de cette manière. Par exemple, il y a des ordinateurs qui changent souvent d'état d'une manière qui n'est pas déterminée par leurs états antérieurs, or leurs réponses (« *output* ») ne sont pas aléatoires de la même manière qu'une table de nombre aléatoires. Que veulent dire les mots « hasard » et « chance » lorsqu'ils s'appliquent à des événements simples ? Van Inwagen³² envisage deux versions de l'argument : soit l'argument s'appuie simplement sur l'idée qu'un événement *indéterminé* relève nécessairement du *hasard*, soit il se fonde sur l'idée qu'un événement, pour être indéterminé, doit être *sans cause*. Dans sa première version, l'argument est peu convaincant, car il n'est pas évident que tout ce qui est indéterminé relève du hasard. Dans sa deuxième version, l'argument est également discutable, car un événement n'a pas besoin d'être sans cause pour être indéterminé. Rien ne prouve que toutes les relations de cause à effet soient nécessaires³³. Il pourrait donc y avoir des relations causales indéterministes, et un événement pourrait être *causé* sans être *déterminé* causalement. Enfin, une troisième conception du hasard, que Van Inwagen ne mentionne pas, définit le hasard ou la chance par le *manque de contrôle*. Or comme l'écrit Kane, il n'y a aucune raison a priori pour supposer que les processus

³² Peter Van Inwagen, *An Essay on Free Will*, p. 129.

³³ Cf. *supra*, p. 302 et s.

indéterminés soient nécessairement *incontrôlés* : « l'indétermination dont il s'agit signifie que le résultat qui advient n'est pas garanti, et non qu'il n'est pas notre action »³⁴.

Dans son article intitulé « The Mystery of metaphysical freedom », Van Inwagen admet finalement que si l'issue d'un choix ou d'un processus n'est pas déterminée, elle est simplement le fait du hasard³⁵. Pour développer cette idée, il imagine Dieu reproduisant la même série d'événements jusqu'à l'instant où l'agent se détermine librement par son choix ou sa décision. C'est l'argument du *replay*, ainsi nommé parce que l'agent y rejoue plusieurs fois la même scène. Cette expérience de pensée permet de voir comment le mot « hasard » peut s'appliquer à une action simple : finalement, selon Van Inwagen, il s'y appliquerait de la même façon qu'à une séquence d'événements³⁶. Les libertariens croient que la liberté suppose une certaine contingence liée au fait que l'action libre n'est pas prédéterminée. Selon cette idée, l'agent n'est libre que s'il peut agir de plusieurs façons différentes dans des situations exactement identiques (la série des événements passés restant la même). Donc si l'agent remontait dans le temps jusqu'à l'instant du choix, il pourrait se comporter autrement que la première fois.

Imaginons une jeune fille, Alice, dans une situation où elle a le choix entre mentir et dire la vérité. Elle ment, mais Dieu lui donne une deuxième chance en la plaçant à l'instant où elle était juste avant de mentir. Puisque son choix n'est pas plus déterminé que la première fois, elle peut tout à fait dire la vérité. Imaginons que Dieu la remplace 700 fois au même moment. Elle ment 500 fois, et dit la vérité 200 fois. Si après cette série de *replays*, elle revenait encore avant l'instant du choix, nous serions tentés d'utiliser ces statistiques pour calculer la probabilité du résultat : il y aurait a priori cinq chances sur sept pour qu'Alice mente, et seulement deux pour qu'elle dise la vérité. Or ce calcul n'est pas correct, puisque par hypothèse, chaque *replay* est indépendant des autres ; les *replays* antérieurs n'ont aucune influence causale (ni logique) sur le dernier. A chaque nouveau *replay*, Alice est dans la même indépendance causale que la première fois, puisqu'elle possède toujours la même liberté. La probabilité du mensonge reste donc à chaque fois la même (par exemple, $\frac{1}{2}$). Il ne

³⁴ R. Kane, *The Significance of Free Will*, p. 131.

³⁵ « *merely a matter of chance* ».

³⁶ Van Inwagen neutralise ainsi la première réponse qu'il a donnée au *Mind argument* dans *An Essay on Free Will* (p. 128).

faut pas raisonner ici comme aux jeux de hasard, où les probabilités ont un fondement réel. Si en 1000 lancers d'un même dé, on obtient 200 fois le 6, la conclusion la plus naturelle de cette observation est que le dé est pipé. On suppose qu'une de ses propriétés physiques rend la probabilité du six supérieure à un sixième ; par exemple, il peut contenir un petit morceau de plomb qui le déséquilibre au bénéfice de son utilisateur (c'est là un procédé classique), et pour voir si le dé est effectivement pipé, on le passe aux rayons x. Dans le cas du *replay*, la situation est différente, puisque par hypothèse, Alice n'a aucune propriété qui rende la probabilité d'un résultat supérieure à l'autre. Donc s'ils ne se produisent pas avec la même fréquence, il ne faut pas croire que leur fréquence relative donne une indication, même approximative, sur la probabilité du résultat dans le prochain *replay*. Au jeu de dé, le hasard apparent n'est pas une indétermination réelle, et si nous parlons de « hasard » dans ce genre de cas, c'est parce que nous sommes incapables de prévoir les événements avec certitude. Dans le cas d'Alice, au contraire, nous avons affaire à quelque chose de profondément mystérieux : un hasard *réel*, si l'on peut dire, c'est-à-dire objectif et sans rapport avec les limites de notre connaissance, si bien que même un observateur omniscient comme le démon de Laplace ne pourrait prédire son choix. Mais dans les deux cas, le mot « hasard » signifie que nous ne contrôlons pas le résultat, et c'est là le nerf de l'argument tel qu'il est formulé par Van Inwagen : si le choix d'Alice est dû au hasard, comment soutenir qu'elle a choisi *librement*, et qu'elle est moralement responsable de son choix ? Si elle choisit bien, elle n'a aucun mérite, mais de la *chance*.

Van Inwagen trouve cette objection très convaincante, et conclut que le libre arbitre est une faculté profondément mystérieuse. Pour ma part, je ne suis pas aussi sensible à la force intuitive de cette objection. La notion de hasard est équivoque, et comme je viens de le suggérer, elle n'implique pas nécessairement l'absence de cause. Par exemple, si je joue à lancer des fléchettes dans une cible en carton, le fait que j'atteigne le cœur de la cible est dû à mon effort de précision, mais aussi à d'autres causes que je ne contrôle pas et dont je ne saurais prévoir les effets. L'action de telles causes est un hasard de mon point de vue, dans la mesure où je ne la contrôle pas et ne peux pas la prévoir. Ce qui vient du hasard vient de causes que je ne contrôle pas, et c'est la raison pour laquelle je n'en suis pas responsable. Or dans le cas d'Alice, le hasard n'est pas de ce genre, car Alice est toujours la seule cause de l'événement, ou au moins la cause principale. La contingence de l'événement ne se ramène

pas à l'imprévisibilité, mais à l'absence de causes *autres* que l'agent lui-même. Il est possible que l'argument du *replay* repose sur une confusion entre ces deux types d'indétermination. Pour éviter toute confusion, il faudrait renoncer à employer le mot « hasard » ou le mot « chance » à propos de la situation où se trouve Alice. Cette distinction faite, que vaut l'argument du *replay* ?

D'abord, il faut distinguer l'indétermination liée aux causes extérieures, et l'indétermination liée à l'agent lui-même (au fait que ses actions ne soit pas causées)³⁷. Le fait que le résultat soit incertain ne signifie pas que l'agent n'en soit pas la cause, car cette incertitude peut tenir à l'agent lui-même. Enfin, ces deux types d'incertitudes peuvent s'additionner. Par exemple, le fait que ma flèche ne touche pas le cœur de la cible peut venir à la fois de ma volonté et de causes extérieures : je n'ai pas fait l'effort de concentration nécessaire, et j'ai été déconcentré par une boutade de mon adversaire.

En résumé, il est important de distinguer deux couples de concepts qui ne se superposent pas : causé/incausé, et prévisible/imprévisible. Comme nous l'avons vu, il est possible d'évaluer la probabilité d'événements dont l'existence est entièrement prédéterminée mais dont on connaît mal les vraies causes. Inversement, il est possible de prédire avec certitude (ou de façon quasi certaine) des actions indéterminées quant aux causes mais non quant aux raisons. Imaginons un être parfaitement rationnel comme Dieu, confronté à des situations où il est évident qu'une des solutions est meilleure que toutes les autres. On peut prévoir avec certitude qu'il choisira la bonne solution. En revanche, s'il est confronté plusieurs fois à des situations où les possibilités alternatives sont indifférentes, c'est-à-dire où

³⁷ R. Kane a bien insisté sur cette distinction. cf. Kane, « Responsibility, Luck and Chance : Reflections on Free Will and Indeterminism », *Journal of Philosophy*, 96/5 (1999), p. 217-240, repris dans l'anthologie de Gary Watson (2003). Il y répond à un argument utilisé par Alfred Mele et qui est très proche de celui du *replay*. Mele propose d'imaginer un monde possible voisin du monde actuel et dans lequel l'agent, par exemple un assassin, agit autrement alors que son passé est exactement le même que dans le monde actuel. Le fait qu'il commette le crime dans le monde actuel et non dans ce monde possible semble être une simple question de chance ou de malchance. Mele compare cet exemple à d'autres comme celui-ci : dans le monde actuel, un homme en colère contre sa femme frappe sur une table et parvient à la casser, alors que dans un monde possible voisin du monde actuel, où il frappe également la table, elle ne se casse pas. Ici, on s'accorde pour dire qu'il n'a pas eu de chance, car la table aurait pu ne pas se casser. Mais, comme le fait observer Kane, le cas précédent est différent : dans l'exemple de l'homme en colère, l'indéterminisme est à l'extérieur de l'agent, alors que dans l'exemple de l'assassin, l'indéterminisme est dans l'agent lui-même. Pour Kane, il s'agit d'un *effort indéterminé*. On objecte parfois que pour les libertariens, le choix ou l'action pourrait être différent même si les efforts de l'agent ont été les mêmes jusqu'à l'instant du choix, ce qui suggère que l'issue est une affaire de chance. Mais Kane rejette cette objection, car dans sa théorie, c'est l'effort lui-même qui est indéterminé ; on ne peut donc pas séparer l'effort et le processus indéterministe (il n'y a pas d'abord un effort, puis un processus indéterministe aboutissant à un choix ou au choix contraire). cf. p. 310 et s. dans l'anthologie de Gary Watson (2003).

aucune n'est meilleure que l'autre, dans ce cas l'issue est indéterminée. On peut tout de même établir des statistiques indiquant le pourcentage de situations où l'individu choisit chaque possibilité, puis calculer la probabilité de ses choix futurs dans des situations analogues. Mais il est évident que ce calcul n'aurait aucun sens : Dieu n'étant pas du tout incliné vers un choix plutôt que vers l'autre, ce calcul ne servirait à rien. Dans un cas comme celui-ci, l'incertitude ne signifie pas que Dieu ne soit pas la seule cause de son choix. Il en est bien la seule cause. Ainsi, un événement *indéterminé* peut tout à fait être le résultat de l'agent *seul*, donc être entièrement en son contrôle. Certains objecteront sans doute que c'est une fiction, parce qu'ils ne croient pas à l'existence d'un être entièrement rationnel. Mais alors prenons-la comme une expérience de pensée. Elle nous montre qu'il n'y a pas de lien *essentiel* entre l'indéterminisme et l'absence de contrôle par l'agent. La notion d'incertitude ou d'indétermination n'implique pas que le résultat échappe, même partiellement, au contrôle de l'agent.

L'objection du *replay* ne paraît donc pas décisive. Mais ce n'est qu'une version parmi d'autres du *Mind argument*. Dans la prochaine section, j'aborderai une deuxième version de l'argument, qui oppose l'indéterminisme au contrôle de soi. Ensuite (dans la section 7.9), je présenterai deux difficultés qui concernent d'une façon particulière les théories de la causalité *événementielle* (notamment une troisième version du *Mind argument*). Pour les éviter, il me semble que la meilleure solution pour les libertariens est d'adopter une théorie de la causalité de l'agent. Je consacrerai donc les dernières sections à la causalité de l'agent.

7.8 Le *Mind argument* : le contrôle de soi

Dans les pages qui précèdent, j'ai discuté une première version du *Mind argument*, l'objection du *replay*, fondée sur un lien apparent entre les notions de hasard et d'indéterminisme. Cette objection ne semble pas décisive, car nous pouvons y répondre en soutenant que ce qui est indéterminé peut être causé ou contrôlé par l'agent. Or une autre version du *Mind argument* se fonde justement sur un lien entre le fait de *contrôler* et le fait de *déterminer causalement*. La liberté et la responsabilité supposent que l'agent exerce un contrôle sur ses propres actions et sur les conséquences de ses actions, c'est pourquoi elles ne consistent pas en un processus indéterminé, mais plutôt en une sorte d'*autodétermination*. Imaginons qu'il y ait des processus indéterminés dans notre corps, et que les mouvements de nos membres varient en fonction des résultats de ces processus. Nous ne pourrions alors

contrôler nos mouvements. Pour cela, il faudrait maîtriser les processus physiques indéterminés dont ils sont les conséquences, mais alors ces processus ne seraient plus indéterminés, ou du moins, pas au sens que l'on donne à ce mot en se référant à la physique quantique, car ils seraient déterminés *par l'agent*.

Revenons d'abord à l'exemple du voleur et imaginons qu'il soit contrôlé selon un processus indéterministe par un démon³⁸. En appuyant sur les touches d'une sorte de piano électronique, il influence les décisions du voleur, dont le cerveau est relié au piano par un câble. Un processus physique indéterministe, de type quantique par exemple, relie le piano au cerveau du voleur. Si le démon appuie sur un la, un do, ou un fa, la décision du voleur de ne pas voler l'argent destiné aux pauvres a seulement une certaine chance de se produire. Or le démon appuie sur un do, et le voleur choisit de ne pas voler l'argent. Dans ce cas, le voleur n'est évidemment pas responsable de son acte, même s'il est indéterminé (la décision du voleur étant *causée* par le démon, mais non *déterminée* par lui).

Maintenant, modifions l'exemple en supposant que le piano fonctionne automatiquement, sans être manipulé par un démon ou un autre agent. A l'intérieur du piano se produisent des phénomènes physiques indéterminés dont le résultat détermine le choix du voleur. Le voleur ne paraît pas plus responsable que dans le cas précédent.

Ensuite, on imagine un processus du même genre que dans le deuxième cas, sauf que l'appareil ne se trouve plus à l'extérieur de l'agent, mais dans son cerveau. On lui greffe un appareil dans lequel se produisent des événements indéterminés, qui déterminent ses choix. Là non plus, le voleur ne semble pas responsable de ce qu'il fait.

Enfin, imaginons un dernier cas, qui diffère du troisième seulement par le fait qu'il n'y a rien d'artificiel dans le cerveau de l'agent : en effet, cette fois, c'est une partie *naturelle* de son cerveau qui est le lieu d'événements physiques indéterminés qui déterminent ses choix. Pourquoi le voleur serait-il responsable dans ce cas, s'il ne l'est pas dans le cas précédent ? Supposons qu'on remplace le cerveau de l'agent par une réplique absolument parfaite fabriquée artificiellement. A moins que le cerveau *d'origine* soit essentiel à l'identité de l'agent, cette opération n'a pas d'incidence sur son libre arbitre. Si cela est vrai, l'opération

³⁸ Je reprends ici les exemples de Van Inwagen dans *An Essay on Free Will* (1983), p. 129-133.

inverse ne le modifie pas non plus, et la différence entre le *naturel* et l'*artificiel* est sans pertinence du point de vue du libre arbitre. La seule différence entre un organe naturel et un organe artificiel qui en est la réplique exacte est une différence d'origine.

Ainsi, l'existence de processus physiques indéterminés à l'intérieur du cerveau de l'agent ne favoriserait nullement le libre arbitre. Au contraire, leur caractère indéterminé empêcherait même l'agent de les contrôler, puisque pour en avoir pleinement le contrôle, l'agent devrait pouvoir les déterminer entièrement lui-même. Lorsque le comportement de l'agent, au lieu d'être déterminé par des causes externes, l'est par des causes internes, il n'en devient pas plus libre ; bien au contraire. Par exemple, si j'ai une réaction allergique, il ne s'agit pas d'une action libre, et ce n'est pas moi qui réagit, mais mon organisme³⁹. De la même façon, les mouvements des particules indéterminées dans mon cerveau ne sont pas nécessairement *mes propres mouvements*. Ou bien devons-nous dire que le cerveau est un organe à part, le noyau de notre personnalité ? Si mes mouvements sont déterminés par une cause située dans ma jambe ou dans mon estomac, il ne s'agit sûrement pas d'actions que j'accomplis librement. En est-il autrement lorsque la cause se trouve à l'intérieur de mon cerveau ?

A première vue, nous rencontrons ici un problème d'identité personnelle. Où est le moi de l'agent, le noyau de son identité personnelle, et s'il est dans une partie de son corps, quelle est cette partie ? Par exemple, est-ce son cerveau tout entier ou une zone de son cerveau ? Il n'est peut-être pas nécessaire de répondre à cette question. Que l'agent s'identifie avec son corps tout entier ou avec une partie de son corps, l'action libre ne se définit pas seulement comme une action engendrée par un processus physique indéterminé survenant dans telle ou telle partie de son corps. Une action *libre* (et non seulement *aléatoire*), doit pouvoir s'expliquer à partir des intentions, des désirs et des croyances de l'agent. La liberté ne réside donc pas entièrement dans de tels processus indéterminés. Nous pourrions dire que l'agent est libre lorsque son action est causée par sa raison ou son esprit, sinon la volonté n'intervient pas à la source de l'acte, mais après ou pas du tout. Or une action libre est essentiellement une action volontaire. C'est pourquoi un phénomène physique qui n'est pas en même temps un mouvement volontaire n'est pas le fait de l'agent lui-même, mais d'autre chose : un organe ou une cause interne.

³⁹ Cf. *supra*, p. 283.

Réponse. Cet argument procède par glissements successifs. Dans les deux premiers exemples, il est clair que le voleur n'est pas responsable de son action, car son choix est causé ou déterminé causalement par des événements qu'il ne contrôle pas. Dans le premier exemple, l'action du démon *cause* le choix du voleur sans le déterminer, mais le voleur n'a aucun pouvoir sur cette relation causale indéterministe. Dans le deuxième exemple, un événement quantique à l'intérieur du piano détermine causalement le choix du voleur. Dans le troisième exemple, un processus du même genre se produit, mais à l'intérieur du cerveau de l'agent, dans lequel on a intégré un appareil fonctionnellement équivalent au piano électronique du second exemple. La différence entre le deuxième et le troisième exemple est la différence entre une cause *interne* et une cause *externe*. Elle ne semble pas pertinente du point de vue du libre arbitre. En revanche, il y a peut-être un abîme entre le troisième et le quatrième exemple⁴⁰. Il est possible que le cerveau de l'agent ou une partie de son cerveau soit essentiel à son identité personnelle. Je ne dis pas que c'est le cas, mais seulement qu'il est difficile d'exclure cette hypothèse. Par exemple, il est possible qu'en enlevant à l'agent une partie de son cerveau où se produisent des événements quantiques qui déterminent ses actions libres, on lui enlève la partie la plus importante de son être, tout le reste étant un ensemble d'organes (au sens où il s'agit d'instruments). La différence entre une partie naturelle et une partie artificielle de l'agent est peut-être essentielle du point de vue de la responsabilité, et pour qu'une action soit libre, il est peut-être suffisant qu'elle vienne d'un événement indéterminé survenant dans une certaine partie *naturelle* de l'agent.

Pour soutenir cette hypothèse, Van Inwagen évoque deux modèles différents de l'action libre. Aucun ne semble incohérent. Le premier admet qu'une décision de l'agent peut venir directement de lui en tant que substance, au lieu d'être causée par un événement indéterminé. Il échappe donc au *Mind argument* tel que je viens de le formuler. Quant au deuxième modèle, il s'accorde avec l'idée que l'action libre vient d'un événement indéterminé dans une certaine partie *naturelle* de l'agent. J'examinerai ces deux modèles dans les sections suivantes. Pour l'instant, je me contenterai de les décrire brièvement pour montrer qu'on ne peut pas les exclure d'emblée. Le premier modèle est celui de la causation immanente, ou « *agent-causation* ». Il a été admis par de nombreux philosophes. Par exemple, les philosophes du Moyen-Age pensent que Dieu est la cause de tout ce qui existe, et qu'il est immuable. Cette

⁴⁰ C'est ce que suggère Van Inwagen dans *An Essay on Free Will*, p. 134, sans trancher la question de manière catégorique.

idée pose un problème, car toutes les causes que nous observons dans la nature modifient quelque chose en elles-mêmes lorsqu'elles produisent un effet extérieur. Les scolastiques ont postulé l'existence d'une autre forme de causalité dans laquelle la cause agit sans provoquer de changement en elle-même. C'est la causalité *immanente*⁴¹. Plus récemment, Roderick Chisholm⁴² et Richard Taylor⁴³ ont développé une idée similaire au sujet de l'action humaine. C'est la théorie de la causalité de l'agent. Selon cette théorie, l'origine ultime d'une action libre n'est pas un événement physique ou mental, mais l'agent lui-même. Contrairement au Dieu des scolastiques, l'agent humain change continuellement, mais selon Chisholm et Taylor, ce ne sont pas des événements ou des changements d'état dans l'agent qui causent son action. Dans une théorie comme celle-ci, la causation d'un événement par l'agent ne suppose pas que cet événement soit causé par un événement antérieur. Il y a des événements que l'agent cause directement en tant que substance. Dans l'exemple du voleur, la décision de ne pas voler l'argent peut être produite directement par le voleur. Ainsi, on ne peut pas dire que cette décision échappe au contrôle de l'agent.

Le deuxième modèle de l'action libre s'inspire de l'idée davidsonienne selon laquelle les actions sont causées par des désirs et des croyances de l'agent⁴⁴. Supposons que lors d'une promenade dans les rues de Rome, j'achète une glace au pamplemousse rose. Mon action s'explique à partir d'un désir et d'une ou plusieurs croyances : je désire manger une glace, et je crois que la glace au pamplemousse rose est une des meilleures dans les glaciers romains. Mon action d'acheter une glace au pamplemousse rose est une action libre si elle a été causée de manière non-déviate par ce désir et cette croyance (et bien sûr, si je n'ai pas choisi sous l'effet d'une menace, d'une contrainte ou d'une manipulation). En admettant ce modèle, on peut imaginer que mon désir et ma croyance causent mon choix de manière indéterministe. Dans cette hypothèse, l'agent *contrôle* ses actions parce qu'elles viennent de ses croyances et de ses désirs. Ainsi, le *Mind argument* ne parvient pas à montrer que l'indéterminisme s'oppose à la liberté.

⁴¹ Cf. par exemple Thomas d'Aquin, *Somme théologique*, I, Q.18, a.3, sol.1.

⁴² Roderick Chisholm, *Person and Object*, (1976) ch. II. Van Inwagen se réfère à Chisholm dans *An Essay on Free Will*, p. 136.

⁴³ Richard Taylor, "Determination and the Theory of Agency" (1958), in Sidney Hook ed, *Determinism and Freedom in the Age of Modern Physics*, New York University Press, New York, 1958.

⁴⁴ Van Inwagen s'inspire de ce modèle dans *An Essay on Free Will*, p. 137.

Dans le second modèle, soit les désirs et croyances de l'agent sont causés ultimement par l'agent en tant que substance (et le second modèle se réduit au premier), soit ils n'ont pas d'autre cause que des événements. Or dans une théorie de l'identité psycho-physique, les croyances et les désirs de l'agent s'identifient à des événements physiques dans son cerveau. Si on admet l'identité psycho-physique, le second modèle est donc ce qu'on appelle un modèle de la « causalité événementielle ». Comme le premier, il échappe au *Mind argument* tel que je l'ai formulé dans cette section, mais il a sur le premier modèle l'avantage de ne pas expliquer les actions libres à partir d'une entité mystérieuse ou d'une relation causale qui n'a jamais été observée dans la nature. Pour cette raison, Van Inwagen préfère ce modèle. Mais ce choix n'est pas nécessairement le meilleur, car le modèle de la causalité de l'agent évite certains problèmes de l'autre théorie. Dans les prochaines sections, je montrerai les difficultés propres à chacun de ces modèles. Pour commencer, je discuterai celui de la causalité événementielle, et puisque Robert Kane a donné à cette théorie une forme particulièrement élaborée, je ferai souvent référence à la théorie de Kane.

Comme nous venons de le voir, le *Mind argument* ne prouve pas que l'existence de l'indéterminisme physique à l'origine des actions soit incompatible avec la liberté. Mais on peut revenir à la charge en soutenant que l'indéterminisme physique ne *favorise* pas la liberté. Si la liberté tient au fait que ce sont *nos* croyances et *nos* désirs qui causent nos actions, pourquoi ne pas se contenter d'un modèle où les croyances et les désirs de l'agent causent ses actions de façon déterministe ? Dans ce modèle, l'agent aurait sans doute un plus grand contrôle de ses actions. De plus, cette idée aurait le mérite de lier la liberté de l'agent avec la régularité de son comportement. Souvent, un individu dont le comportement est instable n'est pas vraiment libre, car il cède à des tentations diverses ou règle ses opinions sur celles d'autrui. La liberté permet d'avancer avec constance vers un but et d'agir de manière cohérente. Or de prime abord, l'idée de Kane, selon laquelle la liberté tient à des processus physiques indéterminés à l'intérieur du cerveau, s'accorde mal avec cette opinion. En effet, où est le lien entre la régularité et la liberté, si la source de la liberté est le hasard aveugle des phénomènes quantiques ?

Pour rendre compte à la fois du caractère indéterminé de l'action libre et de son caractère rationnel ou régulier, Kane élabore une théorie dans laquelle la liberté a deux sources : d'une part, les volitions indéterminées (*self-forming willings*) et d'autre part, la

“structure du Moi” (*Self Network*)⁴⁵. Les volitions indéterminées naissent à la manière des phénomènes quantiques indéterminés. Ce sont elles qui font le côté surprenant de beaucoup d’actions libres, et c’est par elles qu’il y a de la contingence – des possibilités alternatives – dans le comportement humain. Néanmoins, lorsque nous connaissons bien une personne, nous sommes capables de deviner sa réaction la plus probable dans certaines situations familières. Cela ne veut pas dire que nous sommes déterminés, mais que nous avons des dispositions qui nous inclinent vers certaines possibilités plutôt que vers d’autres, et c’est là qu’intervient la “structure du Moi”. Elle est l’ensemble des inclinations plus ou moins stables de l’agent : ses *habitus*, ses dispositions et ses instincts. Nous avons vu que pour Kane, le cerveau est un système chaotique dans lequel les phénomènes indéterminés ont tendance à s’amplifier⁴⁶. La structure de ce système chaotique influe évidemment sur la manière dont ils s’amplifient, et ainsi sur leurs effets. Cette structure est celle du *moi*. C’est elle qui confère une régularité aux comportements de l’agent, même lorsqu’ils ne sont pas entièrement déterminés⁴⁷. Elle influe sur la probabilité des effets macroscopiques ayant leur origine ultime dans des volitions indéterminées (*self-forming willings*).

Cette théorie de Kane a plusieurs mérites. Tout d’abord, elle est naturaliste : elle ne postule en dernière analyse que des entités et des phénomènes physiques déjà connus. Elle s’oppose en cela au dualisme. Si nous ne voulons pas multiplier les types d’entités sans nécessité, l’idée d’une âme immatérielle et imperceptible ne peut être admise que s’il n’y a aucune façon cohérente de montrer la place du libre arbitre dans les phénomènes naturels en se référant exclusivement à ce qui est connu par la science contemporaine. La théorie de Kane a indéniablement cet avantage sur les théories de l’agent qui conduisent au dualisme ou à l’idée, assez mystérieuse, des propriétés émergentes. Ensuite, Kane ne se contente pas de

⁴⁵ Kane, *The Significance of free will*, p. 137-142.

⁴⁶ Cf. *supra*, p. 350.

⁴⁷ Plus précisément, Kane fait l’hypothèse suivante : « T37 : L’impression que certains événements dans le cerveau, comme ceux qui correspondent à nos efforts et à nos choix, sont des choses que nous faisons plutôt que des choses qui arrivent simplement, a son fondement dans la superposition d’ondes synchronisées (ou d’oscillations dans l’activation des neurones) : celles de la structure du moi et celles des événements neuronaux. Cette suggestion, en d’autres mots, est que les événements neuronaux correspondant à nos efforts et à nos choix seraient recouverts par les ondes unifiant la structure du moi. Les événements de l’effort et du choix se produiraient pour ainsi dire « à l’intérieur » de la structure du moi dont les oscillations distinctives se superposeraient à eux. » Ces oscillations, relevant de la structure du moi « contribueraient causalement aux choix, posant un réseau concurrent de raisons par dessus, pour ainsi dire, de telle sorte que A serait choisi pour les raisons R, plutôt que B pour les raisons R’ (ou vice versa) – cela confirmerait la croyance que les efforts et les choix sont notre activité, le produit de notre moi ». (*The Significance of Free Will*, p. 140.)

réduire la liberté à l'indéterminisme physique, mais comme nous l'avons vu, il tient compte des habitus et de la régularité des comportements. Il ne conçoit donc pas l'action libre comme un événement purement aléatoire, entièrement indépendant des actions accomplies par l'agent au cours de son passé proche ou lointain. Celui-ci, en agissant, modifie ou affermit ses propres tendances, qui inclinent sans nécessiter⁴⁸, en modifiant la probabilité des futurs contingents dont il a le pouvoir. Cependant, pour Kane, l'action n'est libre que dans la stricte mesure où elle n'est pas causée ; la responsabilité de l'agent vient du fait qu'il est une cause ultime, incausée. Par conséquent, si la structure du moi ne résultait pas elle-même d'événements physiques indéterminés survenant dans le cerveau de l'agent, elle ne ferait que limiter la liberté de l'agent en diminuant la probabilité de certains choix. C'est parce qu'elle vient en partie d'événements indéterminés que l'agent est responsable de ses tendances.

La théorie de Kane répond sans doute de façon satisfaisante à la plupart des objections scientifiques ou philosophiques contre l'idée que les actions libres ont leur origine dans des événements indéterminés à l'intérieur du cerveau. Cependant, elle doit faire face à deux objections. La première objection est relative aux critères de l'identité personnelle. La deuxième est une version du *Mind Argument* assez proche de celle que je viens de discuter dans les pages précédentes. Elle oppose une nouvelle fois l'indéterminisme au contrôle que l'agent doit exercer sur ses actions pour en être responsable. Je vais maintenant présenter ces deux objections. Ensuite (en 7.10), j'aborderai la théorie de la causalité de l'agent. Dans les deux dernières sections, enfin, je discuterai les principales objections contre cette théorie.

⁴⁸ Je prends ici cette expression en un autre sens que Leibniz.

7.9 Difficultés pour la théorie de la causalité événementielle

La thèse libertarienne se heurte à un mystère, mais ce n'est pas une raison suffisante pour la rejeter. Comme l'écrit Van Inwagen, nous avons le choix entre l'énigmatique et l'inconcevable⁴⁹. La thèse libertarienne est énigmatique, mais rend compte avec cohérence de nos intuitions sur la responsabilité morale, ce que ne peut faire une théorie compatibiliste. Dans les réflexions qui vont suivre, je soutiendrai que le défaut de la thèse libertarienne est seulement son côté mystérieux. La vie est mystérieuse, mais personne ne doute qu'elle existe. De la même façon, nous pouvons croire que le libre arbitre existe en l'homme même si cette faculté est étrange. La théorie de Kane, parce qu'elle repose sur l'indéterminisme quantique dans le cerveau, se heurte à des objections plus radicales. La première objection est que cette théorie suppose une conception de l'identité personnelle difficilement compatible avec nos intuitions. Comme le thème de l'identité personnelle nous entraîne vers un sujet beaucoup trop vaste pour être traité ici, je présenterai brièvement cette difficulté, pour m'intéresser plutôt à la deuxième objection. Cette deuxième objection est encore une forme de l'argument du *Mind*.

Tout d'abord, les théories de la causalité événementielle posent un problème lié aux critères de l'identité personnelle. En effet, s'il se produit un phénomène physique indéterminé à l'intérieur de mon bras, le fait que ce bras soit *mon* bras ne suffit pas pour que ce phénomène soit une action libre, une action dont je suis l'agent. De même, le fait que mon cerveau soit une partie de moi ne suffit pas pour que les phénomènes physiques qui s'y produisent soient des actions dont je suis l'auteur. L'identité dont il s'agit ici n'est pas l'identité de la personne à travers les changements. La question que nous devons nous poser n'est donc pas : qu'est-ce qui fait qu'une personne *x* demeure *x* (par exemple Maxime Lepetit) après un changement profond tel qu'un accident grave qui le défigure et lui fait perdre la mémoire ? Notre question est plutôt la suivante : puisque faire *A* librement, c'est le faire *par soi-même*, qu'est-ce qui définit le « soi » en question ? Visiblement, il ne suffit pas que la cause de mon comportement soit à l'intérieur de mon corps pour que je sois responsable de ce que je fais. Mais alors que faut-il en plus ?

⁴⁹ Peter Van Inwagen, *An Essay on Free Will*, p. 150 : « I must choose between the puzzling and the inconceivable. I choose the puzzling ».

Un dualiste pourrait dire que les actions libres ont leur origine causale dans l'âme de l'agent. Tous les mouvements ayant leur origine dans le corps de l'agent seraient alors involontaires. Mais on peut soutenir qu'un tel dualisme prête le flanc à la même objection que la théorie de Kane. Après tout, l'âme n'est pas l'agent lui-même, mais une partie de l'agent, comme le cerveau ou n'importe quel organe. Il y a pourtant une grande différence entre l'âme et les autres parties, car pour un dualiste, elle est au centre de l'identité personnelle de l'agent ; il ne peut pas la perdre sans cesser d'exister, alors qu'il peut perdre un bras ou une jambe sans cesser d'exister. Un organe vital est un organe qu'on ne peut perdre sans cesser de vivre, mais un organe vital peut être remplacé par une prothèse sans que la personne opérée perde son identité. Au contraire, l'âme, si elle existe, ne peut pas être remplacée sans perte de l'identité personnelle. Si cette différence est vraiment pertinente, nous pouvons nous demander si la permanence du cerveau de l'agent est essentielle à l'identité personnelle à travers le temps. Sans entrer ici dans les détails, nous pouvons observer qu'aucune partie du cerveau ne paraît nécessaire à la continuité personnelle. On a même réussi des opérations dans lesquelles les patients se sont fait enlever la moitié du cerveau. Juste après l'opération, les séquelles sont importantes, mais dans certains cas, l'entraînement ou la rééducation permet à la personne opérée de recouvrer la plupart ou même la totalité de ses facultés, surtout si elle a été opérée avant la puberté⁵⁰. De toute façon, nous estimons qu'elle est encore elle-même, c'est-à-dire la même personne qu'avant l'opération, même si elle a perdu définitivement certaines facultés.

Mais l'objection la plus décisive est ailleurs. Dans une théorie de la causalité événementielle, la liberté de l'agent est liée au fait que son action a pour origine un événement indéterministe. Dans cette perspective, on suppose que l'action soit causée par des croyances et des désirs plus ou moins conscients. Si cela est vrai, nous pouvons envisager deux hypothèses. Appelons BD l'ensemble de désirs et de croyances qui cause l'action, ou plutôt leur réalisation physique dans des événements neuronaux. Soit l'action est causée par BD de manière indéterministe (dans ce cas, BD peut exister sans que l'action se produise). Soit l'action est causée par BD de manière déterministe, mais BD n'est pas causé de manière déterministe. Dans la deuxième hypothèse, l'agent ne peut contrôler son action qu'en

⁵⁰ Voir sur ce point John Eccles, *Évolution du cerveau et création de la conscience*, p. 265 et s.

contrôlant BD, car il ne peut rien au fait que si BD se produit, l'action en résulte. C'est là une application du principe du transfert⁵¹.

Dans la première hypothèse, si l'agent ne contrôle pas BD, il ne peut contrôler son action qu'en influant sur le processus indéterministe conduisant de BD à l'action. Dans les deux hypothèses, l'agent ne peut contrôler l'action qu'en influant sur un processus indéterministe. La différence entre les deux hypothèses est liée seulement au fait que l'agent intervienne avant ou après BD. Pour illustrer ce raisonnement, imaginons un appareil comprenant un bouton et deux voyants lumineux⁵². L'un est vert, et l'autre rouge. Si j'appuie sur ce bouton, une des deux lumières s'allume, mais il m'est impossible de savoir laquelle, car le processus qui provoque l'allumage des lampes est de type indéterministe. Supposons que j'appuie sur le bouton. Même si je contrôle cette action, on ne peut pas dire que je contrôle le résultat. Il me serait plus facile de le contrôler si l'appareil fonctionnait de manière purement déterministe. S'il contenait un mécanisme extrêmement complexe, cela pourrait m'empêcher de prédire le résultat. En ce sens, je n'aurais pas non plus le contrôle du résultat. Mais cette absence de contrôle n'est pas une conséquence logique de la description du mécanisme. Au contraire, si le processus est indéterministe, même le démon de Laplace ne peut en prévoir l'issue. Imaginons maintenant qu'un appareil de ce genre soit branché sur mon cerveau. Si le voyant vert s'allume, je vole l'argent, et si c'est le voyant rouge, je ne vole pas l'argent. J'appuie sur le bouton, le voyant vert s'allume et je pars avec l'argent des pauvres. Dans ce cas, je ne suis pas responsable de mon action, et puisque je ne peux pas influencer sur ce qui se passe après l'allumage du voyant vert, ma seule possibilité pour exercer un contrôle dans l'ensemble du processus est d'influer sur la relation causale indéterministe qui aboutit à l'allumage du voyant. Et si on modifie l'exemple de telle sorte que le processus indéterministe se place après l'allumage du voyant rouge ou vert, la conclusion est la même : je ne peux contrôler le résultat qu'en influant sur ce processus indéterministe. De même, si une action libre A est causée de manière indéterministe par une suite d'événements comprenant BD, l'agent doit pouvoir contrôler son action en influant le processus indéterministe, qu'il se situe avant ou après BD.

⁵¹ Cf. *supra* le principe du transfert causal de non-responsabilité, p. 151.

⁵² Je m'inspire ici de Van Inwagen, *An Essay on Free Will*, p. 142 et s.

Or comment l'agent peut-il exercer cette influence, si nous nous plaçons dans la perspective d'une théorie de la causalité événementielle ? Il faut qu'un processus physique agisse sur la relation causale menant de BD à l'action. Mais si ce processus est purement déterministe, l'agent n'y a aucune part. Pour qu'il puisse y jouer un rôle, il faut que ce soit un processus indéterministe. On tombe ainsi dans une régression à l'infini, car le problème qui se posait précédemment au sujet d'une première relation indéterministe se pose également ici. Pour que l'agent contrôle cette deuxième relation indéterministe, il en faut une troisième, etc. Dans une théorie de la causalité de l'agent, il serait plus facile de résoudre ce problème, car l'agent interviendrait directement dans le processus, et il ne serait pas nécessaire d'imaginer une série d'événements physiques agissant sur la relation indéterministe elle-même pour l'influencer. Dans la théorie de Kane, l'issue des processus indéterministes est influencée par la structure du moi (*Self-network*), mais cela ne donne pas à l'agent un contrôle ultime sur le processus indéterministe, car la structure du moi est elle-même l'effet de processus indéterministes. Ultimement, l'agent ne contrôle pas la structure de son moi. Il ne peut donc contrôler les processus indéterministes qui se produisent en les influençant par le biais de cette structure. Je pense donc que si l'idée de la *Self-network* rend bien compte des inclinations, des vices et des vertus, elle ne résout pas l'objection que je viens de présenter.

7.10 La causalité de l'agent

Face aux difficultés rencontrées par les théories de la causalité événementielle, il peut sembler meilleur d'opter pour une théorie de la causalité de l'agent (*agent-causation*). Il ne suffit pas, en effet, que l'action découle d'un événement indéterminé pour que nous puissions la considérer comme une action libre, il faut encore que l'origine ultime en soit l'agent lui-même. L'événement indéterminé peut survenir dans le corps de l'agent sans être le fait de l'agent lui-même. Ainsi, la réflexion sur la responsabilité morale nous oblige à nous demander comment l'agent *lui-même* peut causer son action, ou si l'on préfère, causer les mouvements physiques appartenant à son action. Au XVIII^{ème} siècle déjà, le philosophe écossais Thomas Reid a développé cette idée⁵³. Plus récemment, elle a été défendue par Richard Taylor⁵⁴,

⁵³ Selon William L. Rowe, cette conception remonte à l'évêque Bramhall, au XVII^{ème} siècle. Elle s'oppose au nécessairement de Hobbes, de Locke, et d'Anthony Collins (1717). Cf. W. L. Rowe, *Two Concepts of Freedom* (1987).

Roderick Chisholm⁵⁵ et par Timothy O'Connor⁵⁶. Cette théorie pose elle aussi plusieurs problèmes. Je vais donc exposer les objections les plus sérieuses contre la théorie de la causalité de l'agent, en essayant à chaque fois d'y répondre au mieux. Je commencerai ici par les objections les moins embarrassantes, avant d'aborder les plus radicales dans les sections suivantes.

Une première objection dit que cette théorie viole le principe de causalité. Pour certains philosophes, c'est une évidence que tout ce qui existe a une cause, ou que tout est soumis à la causalité universelle. Par exemple, Anthony Collins, un ami et disciple de Locke, soutenait que toutes nos actions sont soumises à la nécessité causale, c'est-à-dire que chaque action est déterminée par un ensemble de causes et de circonstances telles que lorsqu'elles sont toutes réunies, aucune autre action n'est possible. Je ne reviens pas ici sur la critique du déterminisme. Il me semble que la seule chose évidente, c'est que tout *événement* a une cause, or il ne s'ensuit pas que *tout* ait une cause⁵⁷. C'est ce que répondait Thomas Reid :

J'admets donc qu'un effet incausé est une contradiction, et qu'un événement incausé est une absurdité. Mais il reste à savoir si une volition non déterminée par des motifs est un événement incausé. Cela, je le nie. La cause de la volition, c'est l'homme qui la voulait.⁵⁸

L'idée d'un agent indéterminé est moins difficile à admettre que celle d'un événement indéterminé, car si un événement indéterminé a une cause, il n'en reste pas moins que certaines caractéristiques de cet événement n'ont ni cause, ni explication, comme si elles surgissaient du néant. Au contraire, Reid peut dire que toutes les propriétés d'une action libre sont causées par l'agent lui-même. Quant à l'agent, son existence et ses propriétés sont d'abord l'effet de causes extérieures qui se trouvent dans ses parents. En admettant que chaque action soit causée par l'agent, une question qui se pose alors est de savoir si le fait de causer une action est lui-même une action. Si c'est le cas, pour accomplir l'action A (beurrer une tartine

⁵⁴ Richard Taylor, "Determination and the Theory of Agency" (1958), in Sidney Hook ed, *Determinism and Freedom in the Age of Modern Physics*, New York University Press, New York, 1958.

⁵⁵ Roderick Chisholm, *Person and Object* (1976), ch. II, p. 53 et s.

⁵⁶ Timothy O' Connor, *Persons and Causes* (2000), ch. III, p. 43 et s. La théorie de la causalité de l'agent a également été adoptée sous des formes diverses par Richard Taylor (1958), Randolph Clarke (1993), William Rowe (1987), et William Hasker (1999).

⁵⁷ Sur le principe "tout a une cause", cf. *supra*, p. 326.

⁵⁸ *Letter to Dr James Gregory*, 1793, in Hamilton, *Works*, p. 87.

de pain), l'agent doit accomplir l'action B (causer l'action A), etc. Chaque action supposerait donc une infinité d'actions, ce qui est impossible.

À première vue, cette objection suppose que B soit distincte de l'action A. Si nous disons que B n'est pas une action supplémentaire, la difficulté disparaît. Or rien n'empêche de considérer la distinction entre A et B comme purement conceptuelle. Dans cette hypothèse, « x fait A » implique nécessairement « x fait B », et « x fait B » implique à son tour « x fait C » mais toutes ces propositions ne décrivent qu'une seule action. En effet, le lien entre ces trois faits n'est pas causal mais purement conceptuel. L'événement « x fait C » n'est pas la *cause* de l'événement « x fait A ». Ces événements sont parfaitement simultanés et la relation d'implication qui les unit n'a visiblement aucun rapport avec une relation de causalité.

D'ailleurs, si cette objection était valable, elle s'appliquerait tout aussi bien à la causalité événementielle. Soit deux événements E et F, si E cause F, on peut supposer qu'il cause en même temps et *ipso facto* l'événement « E cause F », que nous pouvons appeler « G ». De la même façon, en causant G, l'événement E cause en même temps l'événement « E cause G », et ainsi de suite, mais ces événements ne sont pas réellement distincts.

On peut néanmoins se demander s'il est correct de dire que l'agent *cause* son action. Comme l'écrit Suarez, « si nous comprenons par le mot “effet” non seulement la chose produite, mais aussi tout ce qui découle du pouvoir de l'agent, alors nous pouvons admettre que l'action est d'une certaine manière un effet de l'agent »⁵⁹. Ainsi, en causant l'événement E, l'agent cause également, en un sens, l'action de causer E. Précisons néanmoins qu'en essayant de faire A, l'agent n'essaie pas *ipso facto* d'essayer A. Par exemple, en essayant de toucher une cible à la carabine à plombs, je n'essaie pas en même temps d'essayer cela. Il faut donc distinguer le fait de *causer par une action volontaire*, et le fait de *causer volontairement*. En outre, si nous voulons éviter d'employer un vocabulaire causaliste ou volitionniste, nous devons éviter de dire que l'agent *cause* son action. En effet, son action n'est pas une suite d'événements causés par un événement interne (le choix ou la « volition »). L'agent est en un

⁵⁹ « Quod si nomine effectus comprehendamus non solum rem productam, sed quidquid a virtute agentis manat, sic concedimus actionem esse aliquo modo effectum agentis » (Francisco Suarez, *Disputationes metaphysicae*, Disputatio XVIII, section X, paragraphe 8 (vol. XXV, p.682 dans les *Opera omnia*, éd. Louis Vivès, Paris, 1861). Cette phrase est citée par Roderick Chisholm dans *Person and Object*, p. 71-72.

sens la *source* ou l'origine causale de tous les événements qui constituent son action, mais ces événements ne sont pas eux-mêmes des actions. Arrêtons-nous un moment sur ce point.

Contre Thomas Reid, Jonathan Edwards⁶⁰ soutient que l'agent ne peut pas causer son acte de volonté sans exercer son pouvoir de produire un tel acte, et sans que cet exercice soit une action distincte. Si cela était vrai, la causalité de l'agent supposerait une régression à l'infini. Pour répondre à cette objection, il semble qu'il faille nier la prémisse en déclarant qu'il y a des actions “basiques”, c'est-à-dire causées par l'agent lui-même sans aucun intermédiaire⁶¹. Lever son bras est souvent une action basique. En un sens, ce mouvement n'est pas directement causé par l'agent, mais suppose l'intervention de causes secondes telles que la contraction de plusieurs muscles et la transmission de signaux nerveux. Cependant ces processus ne sont pas des actions de l'agent ; en effet, ils ne sont pas des choses que l'on peut choisir de faire, mais des événements qui se produisent nécessairement si l'on choisit de faire certaines choses (comme lever son bras).

L'action basique se caractérise par le fait qu'elle n'est pas le but d'une autre action plus élémentaire. Par exemple, si je lève mon bras pour communiquer un message en morse, j'accomplis là une action, et il n'y a aucune autre action que j'accomplisse pour faire ce mouvement du bras. Néanmoins, lorsque j'accomplis cette action, mes muscles sont en activité de telle manière que je lève effectivement mon bras. Notons que des personnes différentes peuvent ne pas avoir le même “répertoire” d'actions basiques, et que pour un même individu, ce qui est une action basique à un certain moment peut cesser de l'être à un autre moment. Par exemple, lever son bras est normalement une action basique, mais si j'ai un problème moteur à la suite d'un accident de voiture, je peux être amené à réapprendre ce mouvement en m'efforçant de contracter tel ou tel muscle.

Mais cette considération sur la nature des actions ne répond pas entièrement aux objections contre la causalité de l'agent, puisqu'elle admet la possibilité que les actions basiques elles-mêmes soient le résultat d'événements physiques complexes au lieu d'émaner directement de l'agent. Si même les actions basiques ne viennent pas directement de lui,

⁶⁰ Jonathan Edwards, *Freedom of the Will*, p. 175-176.

⁶¹ Sur les actions basiques, cf. Roderick Chisholm, *Persons and Causes*, p. 84-87.

pensera-t-on, il n'y a pas d'actions basiques au sens le plus strict, mais toute action procède d'événements, et il n'y a rien dont l'agent soit lui-même la cause.

Je pense qu'il faut répondre ainsi : rien n'empêche de penser que les actions basiques telles qu'on les définit habituellement sont le résultat de processus causés par l'agent et qui ne sont pas eux-mêmes des actions. Ces processus, ou le premier d'entre eux (par exemple, un influx nerveux), pourraient être causés directement par l'agent. Je supposerai pour la suite que cette hypothèse est valable. Toute la difficulté sera de voir comment cette influence de l'agent peut s'insérer dans la causalité physique.

Comme je l'ai dit plus haut, il est douteux que l'action elle-même soit « causée » par l'agent. Cette manière de s'exprimer est peu satisfaisante, car l'action n'est pas un événement ou un mouvement physique comme un autre : elle n'est pas provoquée par l'agent sans que celui-ci soit actif au moment même où l'événement se produit, et actif *dans cet événement même*. Autrement dit, l'action n'est pas ce qui se passe après que l'agent ait agi ; elle est *ce que fait* l'agent. Il n'en reste pas moins que l'agent cause les *mouvements* qui font partie de son action, et il est donc important de savoir comment il les cause. Contrairement à la plupart des actions, certaines actions basiques telles qu'imaginer, calculer, réfléchir, ou prendre une décision, n'impliquent pas de mouvements corporels. Elles “demeurent” dans l'agent, et ne sont pas observables. C'est la distinction médiévale entre l'*actio transiens in objectum* (c'est-à-dire qui “va” vers un objet) et l'*actio manens in agente* (qui demeure dans l'agent)⁶². Dans les actions basiques qui “demeurent” dans l'agent, la question se pose de savoir si l'agent cause quelque chose, mais dans les actions basiques de la première espèce, il est clair que l'agent cause des événements qui font partie de son action, par exemple, il provoque la transmission de signaux nerveux qui entraînent l'activité de certains muscles. Il faudra donc se demander comment il peut causer de tels mouvements sans être lui-même déterminé à les causer.

Une autre objection classique contre les théories de la causalité de l'agent est celle-ci : l'événement dont l'agent est considéré dans cette théorie comme la cause ultime est plus ou moins ponctuel. Or l'existence de l'agent, au contraire, n'est pas un fait ponctuel. Comment donc pourrait-elle expliquer ce qui se passe à un instant précis ? Par exemple, si l'on voit un individu se diriger vers un distributeur automatique et acheter un paquet de chips, il ne suffit

⁶² Cf. par exemple Thomas d'Aquin, *Somme théologique*, I, Q.18, a.3, sol.1.

pas de dire qu'il a fait ce choix par lui-même, en tant qu'agent, pour expliquer entièrement l'action, car on ne voit pas ainsi pourquoi il fait cela *à ce moment* plutôt qu'une heure avant ou après. De plus, pour que l'explication soit complète, il faut qu'elle montre aussi *pourquoi* il a choisi des chips plutôt qu'une barre de chocolat.

Je pense que ces objections ne sont pas décisives. Elles obligent seulement les tenants de la causalité de l'agent à mettre de l'eau dans leur vin en reconnaissant que l'agent a des relations avec la causalité *événementielle*. Bien qu'il soit une cause d'un genre particulier, il est relié ou intégré au monde de la causalité événementielle. Par exemple, s'il se dirige vers le distributeur à t1, c'est parce qu'il a faim. La sensation de faim est un événement qui pousse l'agent à chercher de la nourriture. Elle n'est pas une cause suffisante de l'événement, mais si l'agent n'a pas faim avant t1, le fait qu'il ait faim à t1 explique pourquoi il se dirige à ce moment vers le distributeur. Chisholm et les autres défenseurs de la causalité de l'agent⁶³ prennent soin de montrer comment l'agent peut subir l'influence d'événements sans cesser d'être la cause ultime, ou au moins la cause principale.

On ne peut donc pas reprocher aux théories de la causalité de l'agent de négliger l'influence que subit un homme dans l'usage de son libre arbitre, par exemple lorsqu'il écoute l'opinion ou le conseil d'autrui. L'expression d'un conseil ou d'une opinion fait partie des événements qui influencent l'agent dans ses choix. Imaginons comme dans l'exemple précédent, un individu qui hésite entre un paquet de chips et une barre de chocolat. Cette fois, il est en compagnie d'un ami qui lui conseille la barre de chocolat, et il suit ce conseil. Ici, le choix (libre) de l'agent est *influencé* si le conseil de son ami "l'incline sans le nécessiter", c'est-à-dire s'il ne le rend pas nécessaire mais probable. On dit qu'un facteur *f* influence le choix *C* si la présence de *f* augmente la probabilité de *C* (donc si en l'absence de *f*, la probabilité de *C* aurait été plus faible).

⁶³ Cf. par exemple Timothy O'Connor, *Indeterminism and Free Will*, p. 187 et s.

Roderick Chisholm⁶⁴ évoque d'autres manières d'influencer un choix sans le causer :

- On peut agir sur quelqu'un sans déterminer son choix jusque dans le détail. Par exemple, on peut contraindre un individu à quitter un lieu sans déterminer le lieu où il ira.
- Si f est une condition nécessaire de C , on peut contribuer à C en produisant f et en ne produisant pas les conditions nécessaires à d'autres choix, voire en empêchant ces conditions nécessaires de se réaliser. Par exemple, en empêchant un individu de lire Nietzsche et d'en entendre parler, on l'empêche de devenir nietzschéen ; et en empêchant un enfant de voir une boîte de chocolat, on évite qu'il ait le désir d'en manger. On peut aussi influencer quelqu'un en lui fournissant les moyens utiles ou nécessaires pour réaliser une action, par exemple, on peut inciter quelqu'un à accepter une invitation en lui donnant un moyen de transport ou en lui remboursant le prix du déplacement.

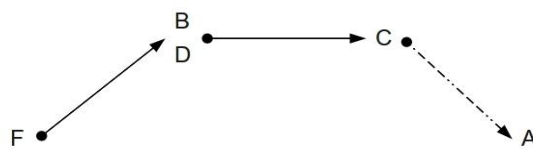
En répondant à ces objections, j'ai voulu montrer que l'hypothèse de la causalité de l'agent n'entraîne pas de contradiction ni de difficulté conceptuelle insurmontable, du moins si nous supposons que l'agent s'intègre dans la causalité événementielle. Dans les prochaines sections, j'exposerai deux objections contre les théories de la causalité de l'agent. La première est le *Mind argument* dans la version que j'ai déjà exposée en 7.9. J'ai montré que cette objection met en difficulté les théories de la causalité événementielle. Je soutiendrai maintenant que cette objection a moins de force contre les théories de la causalité de l'agent. La deuxième objection, plus forte à mon avis, est qu'il est impossible de concilier la causalité de l'agent avec le matérialisme et le naturalisme.

⁶⁴ Cf. Roderick Chisholm, *Person and Object*, p. 67-69.

7.11 L'indéterminisme s'oppose-t-il au contrôle de soi ?

Les libertariens soutiennent que le déterminisme exclut le libre arbitre, l'exercice de la liberté supposant une certaine absence de déterminisme à l'intérieur de l'agent. On dit parfois que la thèse libertarienne suppose une rupture dans la causalité des processus qui sont à l'origine de l'action libre, mais il n'est pas nécessaire d'aller aussi loin. Nous pouvons nous contenter de dire qu'il y a des relations causales non déterministes en quelque point de ces processus. Après tout, comme le note Clarke, « un processus causal indéterministe peut inclure des relations causales exactement aux mêmes endroits que dans un processus entièrement déterministe »⁶⁵. Ainsi, tout ce qui arrive dans l'agent s'intègre parfaitement dans la causalité.

Une question aussi importante que délicate est donc de savoir en quoi consiste cet indéterminisme et où il se situe exactement : par exemple, entre le choix et l'action, ou entre le choix et les raisons d'agir ? J'ai discuté plus haut l'idée que l'indéterminisme n'augmente pas la responsabilité morale mais ne peut au contraire que la diminuer en limitant le contrôle que l'agent a de ses choix, de ses actes, ou de ses états mentaux. Si l'indéterminisme se situe entre le choix (ou le vouloir) et l'action dont l'agent est censé être responsable, il ne semble en effet que diminuer le contrôle exercé par l'agent sur ses propres actions. Nous pouvons représenter ainsi cette première hypothèse :



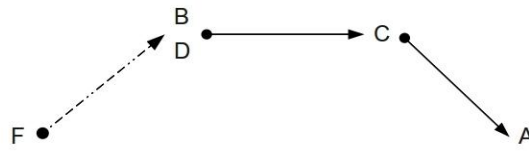
B : croyance ; D : désir ; C : choix ou volonté ; A : action ; F : causes ou facteurs que l'agent ne contrôle pas

flèche en pointillés : relation causale indéterministe. Flèches continues : relations déterministes.

⁶⁵ Randolph Clarke, *Libertarian Accounts of Free Will*, p. 73.

Prenons un exemple. Jean met de la musique. Il s'agit du *Pierrot lunaire* de Schönberg. Marie proteste aussitôt, car elle trouve cette musique angoissante et voudrait plutôt écouter de la musique baroque. Mais Jean n'a pas voulu contrarier Marie : il a lancé la lecture aléatoire, et avait peu de chances de tomber sur cette œuvre. En dehors de cette œuvre moderne, tout le répertoire du lecteur est baroque. La responsabilité de Jean est moins grande que si le processus n'était pas aléatoire. S'il prenait le disque du *Pierrot lunaire* et appuyait sur le bouton “*play*”, il serait clairement responsable, car il saurait qu'en faisant cela, il ferait *nécessairement* entendre cette musique. Le déterminisme entre C et A augmente la responsabilité au lieu de l'exclure. Donc si la liberté suppose un processus indéterminé, il ne peut pas être entre C et A. Il doit plutôt se situer avant C.

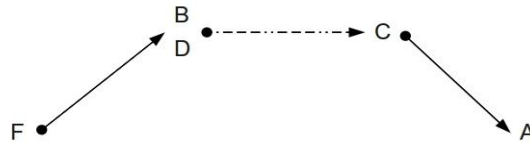
Imaginons d'abord qu'il se situe entre F et BD. Le choix semble être une conséquence des croyances et des désirs de l'agent. Il paraît donc assez naturel de penser que les croyances et les désirs de l'agent déterminent ses choix. Dans le cas d'un conflit entre deux désirs de l'agent, le plus fort l'emporterait et provoquerait nécessairement le choix correspondant. Pour que l'action ne soit pas déterminée par des causes externes ou internes agissant sur l'agent sans qu'il y soit pour rien, il faudrait donc que ses croyances et ses désirs ne soient pas causés de façon déterministe. Imaginons par exemple que Jean mette un disque de Haendel sur le conseil de Marie, un disque qu'on lui a prêté et qu'elle trouve excellent. Si Marie n'avait pas exprimé son opinion, Jean aurait probablement choisi autre chose, par exemple un quatuor de Brahms. Jean aurait peut-être mis spontanément un disque de Haendel qui se trouvait à côté de l'appareil, mais à partir du moment où Marie l'a conseillé, il est devenu très probable que Jean fasse ce choix. Dans ce cas, Jean est influencé par le conseil de Marie, mais agit librement parce qu'il avait la possibilité d'agir autrement. La relation entre le conseil de Marie et le choix de Jean est une relation probabiliste. Le conseil de Marie a seulement augmenté la probabilité de ce choix. Marie a influencé les désirs de Jean en lui décrivant l'enregistrement ; elle l'a incliné sans le nécessiter.



Bien que ce schéma paraisse acceptable au premier abord, il pose un problème. En effet, pour que Jean soit responsable de son action, il faut qu'il contrôle ce qui en est l'origine, en l'occurrence BD. Or le fait que BD ne soit pas déterminé par F ne suffit pas pour que Jean en soit responsable. Imaginons que BD soit provoqué par un randomiseur connecté au cerveau de Jean. Dans cet appareil se produit un processus quantique indéterminé qui entraîne le désir d'écouter du Haendel, ou celui d'écouter du Brahms. Jean choisit Brahms. Est-il responsable de ce qu'il fait ? Apparemment non, puisqu'il ne contrôle pas le processus indéterministe qui se déroule dans l'appareil. Une fois de plus, nous pouvons utiliser le principe du transfert de non-responsabilité :

1. Jean n'est pas responsable de F.
2. Jean n'est pas responsable du fait que F cause BD.
3. Donc Jean n'est pas responsable de BD.

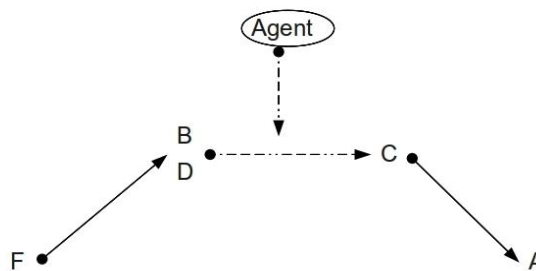
Habituellement, l'argument du transfert est employé contre la thèse déterministe. Ici, dans la deuxième prémisse, on a remplacé la relation causale déterministe par une relation probabiliste, mais le transfert est toujours valable, tant que l'agent n'est pas responsable de cette relation. L'indéterminisme ne suffit donc pas pour que l'agent exerce un contrôle de ses actions ou de ses états. Si ses croyances ou ses désirs dérivent de façon plus ou moins aléatoire d'un facteur qu'il ne contrôle pas, il faut sans doute qu'il exerce une influence sur ce processus indéterministe, par exemple, en augmentant la probabilité d'un événement mental. Nous pouvons appliquer le même raisonnement à l'hypothèse dans laquelle un processus indéterministe aurait lieu entre BD et C, comme dans le schéma ci-après.



En effet, là aussi, nous pouvons appliquer le principe du transfert :

1. Jean n'est pas responsable de BD.
2. Jean n'est pas responsable du fait que BD cause C.
3. Donc Jean n'est pas responsable de C.

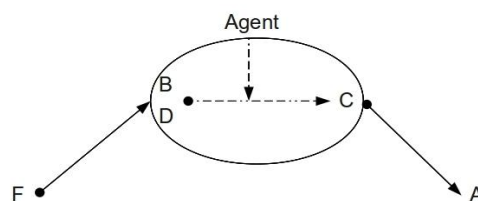
Dans ce schéma comme dans le précédent, l'existence d'un processus indéterministe avant le choix n'est pas une garantie suffisante de la responsabilité morale. Tant qu'il s'agit seulement d'un processus quantique non contrôlé par l'agent, celui-ci n'est pas responsable de son choix ni de son action. Il faudrait donc supposer que la relation indéterministe est contrôlée par l'agent qui a le pouvoir, par exemple, d'augmenter ou de diminuer la probabilité du résultat. En corrigeant le dernier schéma, on peut donc imaginer ceci, où P représente l'agent :



C'est ainsi que Randolph Clarke présente la place de l'agent dans la causalité physique. L'agent contrôle la manière dont les raisons d'agir influencent son comportement :

Supposons qu'un agent humain soit un processus entièrement événementiel (*event-causal*). Si le ciel est nuageux, j'acquies la croyance que le ciel est nuageux et qu'il pourrait pleuvoir. Si je crois que le ciel est nuageux et qu'il pourrait pleuvoir, alors je peux prendre mon parapluie. Mais il est très probable que je ne le prenne pas. Cependant, si je crois que le ciel est nuageux et qu'il pourrait pleuvoir, et si en outre mon compagnon remarque que le ciel est nuageux, alors il est fort probable que je prenne mon parapluie. La remarque de mon compagnon, ou l'absence de cette remarque, peut influencer causalement sur le fait que les nuages et ma croyance causent une certaine action.⁶⁶

Cette description ne convainc pas Timothy O'Connor, parce qu'elle suggère que les raisons (croyances et désirs) sont extérieures à l'agent. Selon lui, il faut plutôt penser que le compagnon en question agit sur l'agent en produisant de nouvelles croyances ou en renforçant des croyances qu'il avait déjà. Cette influence s'ajoute aux désirs et aux croyances qu'il possédait avant. Il faut tenir compte du fait que les croyances et les désirs font partie de l'agent et exercent une pression sur lui. C'est ce qu'on ne voit pas dans ce schéma, car l'agent y est représenté comme une entité séparée de BD. Les croyances et les désirs ont une certaine force indépendamment de l'agent, c'est-à-dire avant qu'il intervienne pour les corriger, les réprimer ou les renforcer. Lorsque des raisons inclinent l'agent à faire A, elles l'inclinent en même temps à ne pas leur résister, si l'on peut dire. Le bon schéma serait donc à peu près celui-ci, où l'agent est représenté par une ellipse :



B, D, et C sont à l'intérieur de l'ellipse parce que ce sont des états ou des événements qui sont dans l'agent lui-même. Les flèches qui partent de B et D signifient que les croyances et les désirs de l'agent exercent une pression sur lui et l'incitent à ne pas résister à la tendance qui

⁶⁶ Randolph Clarke, "Toward a credible Agent-causal Account of Free Will", *Noûs* 27 (1993), p. 201, n. 11.

conduit naturellement de BD à C. L'agent garde pourtant la capacité de contrer cette influence, et donc de ne pas faire A.

Ce schéma n'est sans doute qu'une représentation grossière et simplificatrice. Mais il indique l'essentiel, à savoir que l'indéterminisme doit se situer à l'intérieur de l'agent et être contrôlé par lui, ce qui suppose finalement deux indéterminismes à la fois, du moins si nous croyons qu'il y a un processus quantique standard entre BD et C. En effet, comme nous l'avons vu, l'existence d'un tel processus ne suffit pas. Il faut qu'il soit contrôlé par l'agent et que cette activité d'auto-contrôle soit elle-même indéterminée, ce qui donne une combinaison de deux phénomènes indéterminés.

Mais cette combinaison est-elle nécessaire ? Une autre solution consiste à imaginer qu'il y a un seul phénomène indéterministe : l'activité de l'agent contrôlant C tout en subissant l'influence de BD. Cette hypothèse est sans doute concevable, mais à mon avis, elle exclut l'idée que la relation causale entre BD et C soit de type quantique, puisqu'un processus quantique n'est en lui-même qu'un phénomène physique sans rapport avec la liberté et pouvant exister dans la matière inerte.

Ce que nous venons de voir permet de comprendre comment les théories de la causalité de l'agent peuvent répondre au *Mind argument* et à l'objection du *replay*. Van Inwagen présente le *Mind argument* en imaginant d'abord un individu connecté à un appareil où se produit un événement physique indéterminé⁶⁷. Le choix de l'individu est le résultat de ce processus indéterminé. Il paraît évident que dans ces conditions, l'individu n'est pas responsable de son choix, même si celui-ci ne découle pas d'un processus déterministe. Or si on imagine plutôt que l'appareil est installé à l'intérieur de l'agent, la conclusion ne semble pas différente : il n'est pas responsable de son choix car celui-ci découle d'un processus qu'il ne contrôle pas. Même si le processus se déroule dans le cerveau de l'agent, ce n'est pas suffisant pour qu'il soit responsable du résultat, car nous ne maîtrisons pas tous les processus qui se déroulent dans notre cerveau.

⁶⁷ Peter Van Inwagen, *An Essay on Free Will*, p. 126 et s. Van Inwagen imagine d'abord un démon qui contrôle le cerveau de l'agent en tapant sur les touches d'un clavier. Ensuite, il imagine un appareil automatique fonctionnant de façon indéterministe (un « piano quantique »). Enfin, il imagine qu'un appareil du même genre, c'est-à-dire, indéterministe, se trouve à l'intérieur du cerveau de l'agent. Dans chacune de ces trois situations, selon Van Inwagen, notre intuition reste la même : l'agent ne paraît pas responsable de ce qu'il fait.

Il faut donc que l'agent contrôle au moins partiellement le processus qui aboutit au choix ou à l'action. Or cela est a priori envisageable dans le cadre d'une théorie de la causalité de l'agent. C'est pourquoi le *Mind argument* est beaucoup moins gênant pour ces théories que pour celles de la causalité événementielle. Dans les théories de la causalité événementielle, le processus indéterministe n'est rien de plus qu'un processus quantique standard, mais dans une théorie de la causalité de l'agent, il y a un processus indéterministe qui est l'activité de l'agent lui-même. La difficulté, alors, est plutôt de savoir en quoi consiste une telle activité et comment elle s'insère dans la causalité physique, par exemple si l'agent a une âme qui influe sur certains processus à l'intérieur du cerveau. C'est ce point que je vais aborder maintenant pour terminer. Je me contenterai d'exposer la difficulté pour en donner un aperçu et montrer qu'elle n'est pas a priori insoluble. Mais soutenir une solution argumentée à cette difficulté demanderait une recherche approfondie en philosophie de l'esprit, ce que je ne pourrai faire ici.

7.12 Le libre arbitre et le physicalisme

En abordant le *Mind argument* sous ses différentes formes, nous avons vu que la référence aux phénomènes quantiques n'est pas toujours pertinente pour mieux comprendre la liberté humaine. L'existence de certains phénomènes physiques indéterminés à l'intérieur du cerveau humain ne permet pas à elle seule de comprendre comment l'exercice du libre arbitre peut s'insérer dans la causalité physique. La solution la plus naturelle est de supposer que l'agent lui-même a un certain contrôle de ces processus. Mais que signifie cela ? Est-ce l'âme de l'agent qui influe sur la probabilité de ces événements indéterminés ? Ou bien est-ce l'agent lui-même (un agent purement matériel) qui contrôle ces processus par le biais de propriétés émergentes ? A mon avis, c'est ici qu'apparaît la difficulté la plus importante pour tous ceux qui défendent la position libertarienne. Le principal problème n'est pas lié au caractère plus ou moins inexplicable de l'action libre, mais au fait qu'elle ne peut se réduire à un ou plusieurs événements physiques standard. J'exposerai cette difficulté en utilisant systématiquement le principe du transfert. Je ne proposerai aucune solution à ce problème. Pour cela, il faudrait élaborer une théorie de l'esprit, ce que je ne peux faire ici. Mon intention est donc seulement

d'exposer la difficulté. Je reprendrai d'abord une argumentation de Jan Cover et John O'Leary Hawthorne.⁶⁸

Le matérialisme, appliqué à l'être humain, revient à soutenir ceci :

Affirmation 1 : Les hommes sont des êtres entièrement matériels.

Or la plupart des matérialistes ne croient pas que les états mentaux s'identifient purement et simplement à des états physiques du cerveau. Ils soutiennent plutôt ceci :

Affirmation 2 : Les événements mentaux *surviennent* sur des événements microphysiques.

Rappelons ce qu'est la *survenance* :

(déf) Un ensemble de propriétés A *survient* sur un ensemble de propriétés B si seulement si, lorsque quelque chose a les propriétés B, il a nécessairement les propriétés A.⁶⁹

La même définition s'applique aux autres types d'entités que les propriétés, par exemple aux événements ou aux états de choses. Donc dire que les événements mentaux surviennent sur des événements microphysiques du cerveau signifie que chaque terme ou expression désignant un événement mental dénote aussi un événement microphysique dans le cerveau. Les événements mentaux n'ont pas d'existence indépendamment des événements microphysiques qui se produisent dans le cerveau.

Cette affirmation 2 peut être comprise de plusieurs manières, par exemple :

2A : Les états mentaux *surviennent* sur des états microphysiques de l'individu.

2B : La distribution globale des états mentaux à un moment t *survient* sur la distribution des états microphysiques à ce moment.

⁶⁸ Jan Cover & John O'Leary Hawthorne, "Free Agency and Materialism" (1996), in *Faith, Freedom, and Rationality*, ed. Jeff Jordan and Daniel Howard-Snyder. Rowman & Littlefield Publishers, Inc. Lanham, Maryland, 1996.

⁶⁹ Cf. Jaegwon Kim, *Trois essais sur l'émergence*, p. 9.

Les discussions récentes en philosophie de l'esprit ont favorisé l'externalisme, c'est-à-dire l'idée que les contenus mentaux dépendent du monde extérieur⁷⁰. Selon Jan Cover et John O'Leary Hawthorne, 2A est fausse parce qu'elle néglige cela, et 2B est fausse parce qu'elle néglige le fait que les contenus mentaux d'un individu dépendent de son passé et de l'histoire. Le matérialisme doit donc se référer à une conception plus souple de la survenance des états mentaux. Appliquée au libre arbitre, cette conception pourrait se formuler ainsi :

3 : La propriété d'un agent consistant à prendre une décision libre à *t* survient sur l'histoire des états microphysiques de cet individu jusqu'à *t*.

Cette affirmation semble plausible. On pourrait lui opposer deux objections, mais qui ne paraissent pas convaincantes :

(a) une décision libre dépend en partie du futur.

(b) Une décision libre dépend en partie de l'environnement microphysique de l'agent.

L'objection (a) s'appuie sur une idée peu vraisemblable, car si nos décisions dépendent de ce que nous imaginons sur le futur, elles ne dépendent pas du futur lui-même. Quant à l'objection (b), elle est douteuse. En effet, nos choix dépendent du monde extérieur, mais pas dans tous ses détails microphysiques. Et même s'il est vrai que la réalisation de nos choix dépend d'événements microphysiques dans le monde extérieur, on ne peut pas dire que nos choix eux-mêmes en dépendent.

Retenons donc 3. Cette formule est d'ailleurs compatible avec différentes versions du matérialisme. Les formes de matérialisme les plus réductionnistes sont les moins compatibles avec le libre arbitre. Par exemple, si un matérialiste pense que chaque croyance d'un individu est une suite d'événements physiques dans le cerveau de cet individu, il admet une thèse plus catégorique que 3 parce qu'il nie toute autonomie des événements mentaux par rapport aux événements neuronaux. Si la thèse 3 pose problème, comme nous allons le voir, alors, *a fortiori*, ce matérialisme pose lui aussi problème. Nous pouvons supposer que pour soutenir

⁷⁰ Sur ce thème, cf. par exemple Fissette et Poirier, *Philosophie de l'Esprit, État des lieux* (2000), ch. V. Par exemple, un plongeur qui voit un bryzoaire, animal d'aspect gélatineux, peut penser « ce bryzoaire est dégoûtant ». Le contenu mental ainsi exprimé ne dépend pas seulement de ce qu'il y a « dans la tête » du plongeur, mais aussi du bryzoaire qu'il voit et des processus physiques qui permettent cette perception visuelle.

la thèse libertarienne sans renoncer à une conception matérialiste de l'homme, il faut attribuer aux états mentaux et aux choix une certaine autonomie par rapport aux états physiques du cerveau, or la survenance est compatible avec une certaine autonomie. Il est donc tentant de soutenir que la dépendance des états mentaux par rapport aux états neuronaux relève de la survenance. Un matérialisme plus fort exclut le libre arbitre, à moins d'envisager une solution dans laquelle les événements mentaux tiennent leur autonomie du fait qu'ils s'identifient à des événements physiques *indéterminés*, mais comme nous l'avons vu plus haut⁷¹, l'indéterminisme physique ne suffit pas pour la responsabilité morale. Il faut que cet indéterminisme soit contrôlé par l'agent. C'est pourquoi il est tentant de combiner l'indéterminisme physique avec la causalité de l'agent. Mais cela suppose que les états mentaux aient une certaine autonomie par rapport aux états physiques du cerveau, car c'est à partir de certains états mentaux (ses choix, etc.) que l'agent influe sur les états de son cerveau, puis à partir de là, sur les mouvements de son corps. Du point de vue matérialiste, il faut donc supposer au moins que les états mentaux *surviennent* sur les états neuronaux. Chaque décision libre surviendrait sur l'histoire microphysique de l'agent jusqu'à l'instant de cette décision. (proposition 3).

Or cette idée exclut la responsabilité morale. En effet, nous pouvons supposer ceci :

T1 : Personne n'est moralement responsable du fait que telle chose survienne sur telle autre.

T1 est analogue à l'idée que personne n'est responsable des lois de la nature. Mais comme l'observent Cover, et O'Leary – Hawthorne, “T1 est encore plus solide, car la survenance est une relation modale relevant de la nécessité métaphysique”.

A cela, on ajoute maintenant :

T2 : Personne n'est responsable du détail de ses états microphysiques.

En vertu du principe du transfert de non responsabilité, T1 et T2 impliquent que personne n'est responsable de ses états mentaux, notamment de ses choix libres. Puisque cette conclusion est fausse, nous devons rejeter une des prémisses : T1 ou T2. T1 paraît solide. La

⁷¹ Dans les sections 7.7, 7.8, et surtout 7.11.

meilleure solution serait donc de nier T2 en déclarant que l'agent contrôle en partie ses états microphysiques. On s'oriente ainsi vers l'idée d'une causation “vers le bas” (*top-down causation*) : l'agent influe sur ses états neuronaux, lesquels déterminent ses états mentaux. Cette hypothèse est intéressante, mais semble poser un problème parce qu'elle viole le principe suivant :

Aucun état de choses x ne peut nécessiter (métaphysiquement) un état de choses y si y contribue causalement à x.

Dans l'hypothèse de la survenance, on peut difficilement imaginer que les états neuronaux d'un homme nécessitent un état mental (par exemple un choix) qui modifie ensuite quelque chose dans ses états neuronaux. Ou plutôt, cela se conçoit, mais seulement sous la forme d'une chaîne causale circulaire entièrement déterministe, et dont le point de départ ultime consiste en des événements microphysiques non contrôlés par l'agent. Pour que la chaîne causale puisse venir “d'en-haut”, c'est-à-dire de l'agent lui-même, ou de ses états mentaux, il faut que ceux-ci aient une autonomie considérable par rapport aux états neuronaux. Cela requiert bien plus que la survenance : l'*émergence* des états mentaux ou de certains pouvoirs causaux de l'agent à partir des propriétés physiques de son cerveau. Il faut qu'une certaine configuration d'états microphysiques dans le cerveau de l'agent engendre au moins des *pouvoirs causaux* permettant à l'agent de contrôler des événements qui se produisent dans son cerveau, et que l'exercice de ces pouvoirs causaux par l'agent ne soit pas entièrement déterminé par les événements ou les entités physiques dont ils tiennent leur existence.

Cette solution a été défendue par Timothy O'Connor et par le biologiste John Dupré. L'un et l'autre soutiennent que les propriétés émergentes sont très répandues dans le monde vivant, végétal ou animal. Si cela est, pourquoi ne pas supposer que le libre arbitre est un pouvoir causal émergent ? Selon O'Connor,

étant donné qu'il n'y a rien d'incohérent dans l'émergence d'une propriété causale “ordinaire” capable d'exercer une influence causale irréductible sur l'environnement dans lequel elle est instanciée, on ne voit

vraiment pas pourquoi il n'y aurait pas une sorte de propriété émergente dont la nouveauté consisterait dans sa capacité de rendre son possesseur apte à produire des changements à volonté.⁷²

En un sens, ce pouvoir causal fait de l'être humain une exception dans la nature, puisqu'il est le seul à le posséder, mais toute métaphysique, même matérialiste et réductionniste, est obligée de voir en l'homme un être exceptionnel, au moins par sa complexité. Selon le biologiste John Dupré,

la solution [...] au problème du libre arbitre est simplement de reconnaître que ce qui est problématique dans les êtres humains n'est pas qu'ils sont des exceptions dans un réseau homogène de connexions causales, mais qu'ils sont des concentrations extraordinairement grandes de capacité causale dans un monde qui en contient peu pour réaliser un tel ordre.⁷³

Mais cette idée n'est pas près de faire l'unanimité chez les philosophes, car pour certains, comme Pereboom⁷⁴, il n'y a pas de preuve claire que l'émergence "forte" existe dans les phénomènes chimiques, biologiques ou psychiques. Or le libre arbitre semble supposer une émergence forte, c'est-à-dire dans laquelle l'entité émergente possède une autonomie considérable. Jaegwon Kim⁷⁵ a formulé un argument intéressant contre le matérialisme émergentiste. Cette thèse admet en général que même si un événement, par exemple une action ou une décision, a une explication qui renvoie aux pouvoirs de l'agent ou à ses propriétés émergentes, il a aussi une explication microphysique⁷⁶. Par exemple, s'il y a une explication génétique au fait qu'un bébé a les yeux marron - c'est-à-dire une explication qui renvoie à un pouvoir causal de type génétique - on peut dire en même temps que ce pouvoir est constitué d'un assemblage de pouvoirs causaux microphysiques qui produisent cette propriété d'avoir les yeux marron. Cela pose évidemment un problème : comment les deux explications peuvent-elles être vraies en même temps ? Ou plutôt : comment chacune de ces

⁷² Timothy O'Connor, *Persons and Causes* (2000), p. 121.

⁷³ John Dupré, *The Disorder of Things*, Harvard University Press, Cambridge (Massachusetts), 1993, p. 216.

⁷⁴ Cf. Derk Pereboom, *Living Without Free Will* (2001), p. 74.

⁷⁵ Cf. Jaegwon Kim, *Trois essais sur l'émergence* (2005). Cet argument a été formulé par Kim dans des ouvrages antérieurs comme *Supervenience and Mind*, Cambridge University Press, Cambridge, 1993. Il est repris par Cover, et O'Leary – Hawthorne, ainsi que par Pereboom (*Living Without Free Will*, p. 75-76).

⁷⁶ Cf. par exemple John Searle : "Puisque toutes les caractéristiques de surface du monde sont entièrement causées par et réalisées dans des systèmes de micro-éléments, le comportement des micro-éléments est suffisant pour déterminer tout ce qui arrive. Une telle vision du monde qui va du bas vers le haut n'exclut pas la causalité vers le bas (*top-down causation*) (notre esprit, par exemple, peut agir sur notre corps). Mais la causalité vers le bas n'opère que parce que le niveau le plus élevé est causé par et réalisé dans le niveau le plus bas." (*Minds, Brains, and Science*, Harvard University Press, Cambridge (Massachusetts), 1984, p. 94)

deux causes peut-elle être suffisante sans que l'autre cause se réduise à elle ? Cela suppose que l'effet est *surdéterminé*.

L'explication microphysique ne semble valoir que si les pouvoirs causaux se réduisent entièrement à des phénomènes microphysiques. Pour Kim, chaque *token* de pouvoir causal à un instant *t* est identique à un *token* de pouvoir causal microphysique à *t*. Il pose que :

(Principe de réduction causale) Si la propriété mentale *M* est réalisée dans un système à *t* en vertu d'une base physique *P*, les pouvoirs causaux de *cette instance de M* sont identiques aux pouvoirs causaux de *P*.⁷⁷

Au lieu de “sont identiques à”, Pereboom propose d’écrire “sont constitués par”. Sous cette forme, le principe est plus facile à admettre, car il n'affirme pas l'identité *token-token* entre les pouvoirs émergents et les pouvoirs microphysiques⁷⁸. Cette objection de Kim reprise par Pereboom semble écarter avec succès un certain genre de matérialisme non réductionniste. L'existence du libre arbitre suppose une émergence plus forte que nous venons d'examiner. Du coup, il paraît nécessaire de se tourner vers des hypothèses qui ne sont pas strictement physicalistes. Il me semble qu'une bonne hypothèse doit satisfaire au minimum les trois conditions suivantes :

- Premièrement, elle doit postuler l'existence de processus indéterminés à l'intérieur du cerveau, que ces phénomènes soient ou non de type quantique. Le recours à une théorie de la causalité de l'agent n'a aucun intérêt si les actions et les choix de l'agent sont entièrement déterminés.
- Deuxièmement, ces phénomènes indéterminés ne doivent pas être purement aléatoires. Ils doivent être contrôlés par l'agent, comme nous l'avons vu plus haut.⁷⁹
- Troisièmement, cela suppose que les processus indéterminés qui sont à la source de

⁷⁷ Je reprends l'argumentation de Pereboom, qui vient elle-même de Kim. Cf. Derk Pereboom, *Living Without Free Will*, p.77, et Jaegwon Kim, *Supervenience and Mind*, p. 126 et s.

⁷⁸ Pereboom critique le principe posé par Kim en donnant l'exemple du bateau de Thésée, qui conserve son identité à travers les multiples réparations qu'il subit. Même si toutes les planches du bateau finissent par être remplacées, il s'agit toujours du même bateau. Donc c'est le même token de bateau, même s'il n'est plus constitué des mêmes planches qu'au départ. De la même façon, le fait que les (token de) pouvoirs causaux émergents se réduisent à des (token de) pouvoirs causaux microphysiques, ils ne leurs sont pas strictement identiques selon l'identité token-token.

⁷⁹ Cf. *supra*, p. 344.

nos actions libres ne soient pas épars ou incohérents. Si le point de départ de chaque action libre est un processus indéterministe dans une infime partie du cerveau, cela donne l'impression qu'il n'y a aucun rapport entre deux actions libres du même individu. D'où la tentation de supposer que ces processus physiques, normalement sans rapport entre eux, sont contrôlés à partir d'un même point ou d'un même centre. Cela permet de concevoir une certaine unité dans les événements indéterminés qui sont à la source des actions libres.

Pour le neurobiologiste John Eccles, il y a dans notre cerveau un processus physique indéterminé au niveau des synapses, et l'âme pourrait contrôler suffisamment le cerveau en modifiant la probabilité de ce processus⁸⁰. Le principal reproche que l'on fait généralement à cette théorie est qu'elle suppose une métaphysique dualiste. Elle postule l'existence d'une entité (l'âme) que la science ne peut ni mesurer, ni observer. Il est en effet incongru de proposer une théorie scientifique où intervient une notion aussi lourde de connotations philosophiques. Au lieu de postuler l'existence d'une âme immatérielle, Timothy O'Connor postule l'existence d'un pouvoir causal émergent : la probabilité des événements indéterminés ne serait pas modifiée par une âme immatérielle, mais par le cerveau lui-même grâce à un pouvoir causal émergent. Une troisième théorie intermédiaire entre les deux précédentes est le dualisme émergentiste de William Hasker. Selon Hasker, l'âme est une entité immatérielle qui émerge à partir d'une base matérielle. Elle est comparable à un champ magnétique : bien qu'il émerge à partir d'une structure de fer, il a une existence propre et exerce par lui-même une influence causale⁸¹.

Mais aucune de ces hypothèses ne paraît compatible avec un physicalisme étroit : soit elles postulent une substance inconnue des sciences naturelles, soit elles imaginent un pouvoir causal tout aussi extraordinaire. Ainsi, il y a bien un « mystère du libre arbitre ». Van Inwagen

⁸⁰ Cf. John Eccles, *How the Self Controls Its Brain*, Springer-Verlag (1994), New York (trad. fr. Béatrice Laroche et Jean-Mathieu Luccioni : *Comment la conscience contrôle le cerveau*, Fayard, Paris, 1997).

⁸¹ Un champ magnétique suffisamment intense serait même capable de subsister si l'on coupe le courant de l'électro-aimant : "Il a été montré que la théorie physique impliquait la possibilité qu'un champ magnétique suffisamment intense maintienne sa propre cohésion par gravité, même si on retire l'aimant qui l'a produit. On peut faire une remarque similaire à propos des trous noirs, qui sont devenus récemment un sujet de réflexion prioritaire en cosmologie". (W. Hasker, *In Search of the Soul, Four views on the mind-body problem*, ed. By Joel B. Green & Stuart L. Palmer). Voir aussi W. Hasker, *The Emergent Self*, p. 99-109. Ce passage porte sur le lien entre la liberté, la causalité de l'agent, et l'émergence "forte".

reproche aux théories de la causalité de l'agent d'expliquer *obscurum per obscurius*⁸², mais en choisissant la thèse de la causalité événementielle, il aboutit lui-même à une énigme. Ainsi, c'est bien la thèse libertarienne *sous toutes ses formes* qui aboutit à une énigme. Est-ce une objection décisive contre la thèse libertarienne en général ? C'est du moins une difficulté qui ne peut être résolue sans une vaste recherche en philosophie de l'esprit et une meilleure connaissance du cerveau humain.

⁸² Peter Van Inwagen, *An Essay on Free Will*, p.151.

Conclusion

Puisque toutes les opinions philosophiques sur le libre arbitre et le déterminisme conduisent à des difficultés plus ou moins grandes, nous avons besoin d'un critère pour classer ces difficultés. Le critère le plus important, que j'ai exposé dans le premier chapitre, est le *principe de crédulité* (de Thomas Reid), selon lequel nous devons tenir pour vrai ce qui nous *paraît* vrai. Pour appliquer ce principe à une affirmation quelconque, il faut comparer les preuves soutenant cette affirmation à celles qui soutiennent l'affirmation contraire. La vérité est du côté de l'évidence la plus forte. La conséquence de ce principe est que nous devons faire le maximum pour « sauver » nos intuitions sur la responsabilité morale, de même qu'un physicien doit faire le maximum pour « sauver les phénomènes ». En effet, nos « intuitions » sur la responsabilité de tel agent dans telle situation font partie de l'*évidence* au sens large, c'est-à-dire de ce qui nous *paraît* vrai. Un second principe inséparable du premier est qu'il faut sauver les phénomènes *de manière cohérente*. En proposant une théorie incohérente, on ne sauve pas les phénomènes, puisqu'en violant le principe de non-contradiction, on s'oppose à une des évidences les plus fortes.

Dans le premier chapitre, j'ai défini la thèse que j'allais soutenir dans les chapitres suivants en employant la méthode suggérée Reid, c'est-à-dire en montrant qu'elle « sauve » nos jugements spontanés sur la responsabilité morale des agents humains dans les cas les plus variés. Je n'ai pas voulu supposer une incompatibilité entre le libre arbitre et toute forme de nécessité, car cette idée me semble avoir des conséquences problématiques (elle suppose un lien conceptuel entre la responsabilité morale et l'indifférence). Mon objectif s'est donc limité à ceci : montrer qu'il y a une opposition entre le libre arbitre et la nécessité *causale*. En ce sens, la thèse que j'ai défendue est un incompatibilisme *causal*.

Dans les chapitres II et III, j'ai défendu les arguments les plus directs et les plus répandus contre le compatibilisme. Il se fondent sur le principe des *possibilités alternatives*, et sur celui du *transfert*. Chacun de ces deux principes exprime une « intuition », c'est-à-dire une idée qui nous paraît vraie de prime abord. En ce sens, ces arguments sont très directs. Mais leur évidence peut être affaiblie par des contre-exemples. Dans le chapitre II, j'ai répondu aux trois types d'exemples utilisés contre le principe des possibilités alternatives. Les plus clairs sont les plus simples, or ils ne montrent pas qu'un agent puisse être moralement responsable sans avoir un certain pouvoir des alternatives. Au pire, ce premier type d'exemples nous

oblige seulement à reformuler le principe des possibilités alternatives pour en préciser le sens. On obtient alors une formulation plus rigoureuse du principe. Peut-être s'éloigne-t-elle de l'intuition de départ en tenant compte des exceptions à la première formulation, mais elle implique toujours une incompatibilité entre le déterminisme causal et le libre arbitre. Pour montrer que la responsabilité morale ne suppose aucun pouvoir des alternatives, même au niveau des intentions et de la volonté, les compatibilistes ont dû inventer des cas plus complexes et surtout moins clairs, donc peu convaincants. Je ne reviens pas ici sur le détail de mes arguments, mais maintenant que nous avons une vue d'ensemble des arguments compatibilistes et incompatibilistes, il est intéressant d'observer que la thèse compatibiliste est généralement plus compliquée que la thèse libertarienne : dans ses critères de la responsabilité morale, dans ses réponses aux objections adverses, et dans ses objections contre la thèse adverse.

Dans le chapitre III, j'ai défendu le principe du transfert d'impuissance face aux nombreux contre-exemples. Relativement simples, cette fois, ils peuvent même appartenir à la vie réelle, mais nous contraignent seulement à formuler le principe d'une manière plus rigoureuse. Le principe du transfert perd ainsi en généralité, mais exprime la même intuition qu'au départ : si je suis incapable d'empêcher que x , et que x cause y , je suis incapable d'empêcher y . La nouvelle formulation du principe est explicitement *causale* là où le principe employait d'abord des notions purement logiques comme l'implication matérielle (dans la deuxième prémisses). Mais ce changement ne limite pas la force de l'argument incompatibiliste. Celui-ci visait le déterminisme causal, or la nouvelle formulation du principe du transfert se distingue justement de la première en ce qu'elle serre de plus près la notion de déterminisme causal.

Dans ce débat entre les compatibilistes et les libertariens, tout l'argumentaire se fonde en dernier ressort sur des exemples. On ne peut prouver sa thèse qu'en réfutant la thèse adverse par des contre-exemples. Mais tous les exemples ne se valent pas. Certains manquent de clarté, voire de cohérence, et une situation réelle est plus convaincante qu'un exemple imaginaire ou invraisemblable. A ce propos, il est utile d'observer que les exemples invoqués par les compatibilistes pour réfuter le principe des possibilités alternatives sont presque tous imaginaires, et parfois si complexes qu'on peut se demander s'ils sont cohérents. Une telle question ne se pose pas devant les exemples libertariens : ils sont simples, clairs, et souvent du domaine de la vie réelle.

Dans le chapitre V, j'ai montré pourquoi les exemples d'omissions, s'ils ne sont pas décisifs, limitent considérablement le nombre des solutions compatibilistes. En effet, la responsabilité d'un événement par omission ne suppose pas que l'agent *cause* cet événement au sens habituel du mot : il *cause* l'événement, mais en un sens contrefactuel, et non au sens de la causalité efficiente. Cela suppose que l'agent ait d'une certaine manière la *possibilité* d'agir pour empêcher l'événement. Il a donc un pouvoir des alternatives. Face à cette objection, les théories compatibilistes se défendent en formulant des critères de plus en plus complexes. Les théories les plus simples, que j'ai examinées dans le chapitre IV, ne sont pas à même de répondre à l'objection. Puisqu'il n'est pas possible de réduire les omissions à des actions, il faut procéder comme Fischer ou McIntyre. En effet, d'un point de vue compatibiliste, la possibilité d'agir autrement ne se comprend que de deux manières : selon une analyse *conditionnelle* du mot « pouvoir », ou selon une analyse en termes de *mondes possibles*. McIntyre choisit la première voie, mais est obligée d'emprunter la deuxième pour éviter les objections les plus classiques contre l'analyse conditionnelle de la liberté. De plus, elle doit spécifier que l'agent est moralement responsable seulement s'il se trouve « dans des conditions normales de délibération ». Un tel critère n'a pas de sens clair. Il faudrait donc définir les « conditions normales » pour qu'elles fonctionnent en tant que critère. Des difficultés analogues apparaissent dans la théorie de Fischer. Au lieu de recourir à l'analyse conditionnelle du mot « pouvoir », Fischer se sert uniquement de la notion de « mondes possibles ». Mais comme je l'ai montré dans le chapitre VI, cette notion ne peut servir dans les critères de la responsabilité morale que si le domaine des mondes possibles pertinents est clairement circonscrit. En effet, il y a une limite nette entre la responsabilité morale et l'absence totale de responsabilité. Fischer et Ravizza ne nous indiquent jamais clairement où est la limite entre les mondes possibles *pertinents* et les autres ; ils font dépendre la responsabilité morale d'une disposition : la sensibilité aux raisons. Or il n'y a pas de limite nette entre la sensibilité *faible* et la sensibilité *modérée* aux raisons. Finalement, le concept de sensibilité modérée aux raisons apparaît comme un hybride entre la notion de disposition et le pouvoir des alternatives compris de manière libertarienne. En effet, s'il s'agissait d'une disposition, on pourrait l'attribuer à l'agent sans accorder une importance particulière aux mondes possibles les plus proches du monde actuel. Si nous leur accordons une telle importance, c'est sans doute parce que nous voulons savoir si l'agent pouvait agir autrement dans les conditions mêmes où il se trouvait au moment de l'action.

Dans le chapitre VI, nous avons vu aussi qu'un critère compatibiliste fondé sur la référence à des mondes possibles a le défaut de ne pas tenir compte des causes ou des facteurs qui influencent l'agent *dans le monde actuel*. Un libertarien peut donc critiquer ce genre de critère au moyen d'exemples où l'agent possède la sensibilité aux raisons (ou un certain pouvoir défini de manière conditionnelle) alors qu'il n'est pas responsable, à cause d'un facteur qui agit sur lui. Pour réfuter les critères de la responsabilité morale proposés par les compatibilistes, les libertariens peuvent employer quatre types d'exemples. En effet, la responsabilité peut être exclue par quatre types de facteurs : une cause externe, une cause interne, un défaut externe, ou un défaut interne. Le libertarien peut réfuter un ensemble de critères EC en trouvant des cas où l'agent, bien qu'il remplisse toutes les conditions EC, n'est pas responsable parce qu'il est influencé par une cause externe (comme un savant fou), une cause interne (comme un parasite) ou un défaut interne (comme un handicap mental). Mais pour protéger EC contre ces exemples, le compatibiliste pourra toujours ajouter une condition « fourre-tout » en spécifiant, comme McIntyre, que l'agent doit choisir dans des « conditions normales de délibération », et ainsi tous ces exemples seront pris en compte par le nouvel ensemble de critères. Cependant, cette opération est purement *ad hoc*.

Puisqu'il n'est pas possible de sauver toutes nos intuitions de manière compatibiliste, il faut conclure que la liberté suppose un pouvoir des alternatives. Concrètement, cela implique l'existence de processus indéterminés à la source de nos actions. Dans le chapitre VII, j'ai présenté les deux objections principales contre cette idée. La première consiste à montrer que le monde physique est entièrement déterministe. J'ai soutenu que cette idée n'a pas de fondement conceptuel, ni empirique. La deuxième objection est le *Mind argument* : si nos décisions résultent d'un processus physique indéterminé, ils nous échappent, à moins que nous le contrôlions par le biais d'un autre processus physique. Mais si ce deuxième processus influant sur la probabilité du premier est purement déterministe, nous retombons sur la même difficulté. Cela découle du principe du transfert : supposons que *x* cause mon choix *y* de manière indéterministe, alors si je ne puis empêcher que *x*, ni que *x* cause mon choix (de manière indéterministe), alors je ne puis empêcher que mon choix soit *y*. Pour éviter une régression à l'infini, il est assez naturel de penser que nous contrôlons le processus indéterministe *directement*, en tant qu'*agents*, au lieu de l'influencer par le biais d'un autre processus. Ainsi, on aboutit à une théorie de la causalité de l'agent, dans laquelle certains événements physiques indéterminés sont causés par l'agent lui-même *en tant que substance*. Mais une nouvelle fois, le principe du transfert nous indique une difficulté : cette hypothèse

elle-même n'aurait aucun intérêt si l'agent se réduisait *entièrement* à un ensemble d'événements ou de propriétés physiques, car ces événements ou ces propriétés échappent ultimement à son contrôle. On pourrait donc raisonner de la manière suivante : l'agent ne peut rien au fait qu'il ait ces propriétés, ni à la manière dont elles influencent les processus indéterministes qui ont lieu dans son cerveau, donc il n'a aucun contrôle de ces processus indéterministes. Ainsi, il semble que l'agent ne puisse contrôler ses actions s'il n'a des propriétés mentales ou des pouvoirs causaux *irréductibles* à ses propriétés physiques ou aux pouvoirs causaux de son cerveau. On peut donc se demander si le *matérialisme* est compatible avec l'idée de responsabilité morale.

Mais cette question est un défi lancé à la philosophie de l'esprit, plutôt qu'une objection contre la thèse incompatibiliste. En effet, elle ne se poserait pas si le principe du transfert n'était pas correct. Or s'il est correct, la thèse incompatibiliste s'impose à nous. Puisqu'il n'est pas douteux, il ne nous laisse que deux possibilités : la thèse libertarienne, ou l'incompatibilisme dur, qui rejette la croyance au libre arbitre et à la responsabilité morale. Mais puisqu'il n'y a aucune raison évidente en faveur du déterminisme, l'incompatibilisme dur a un coût beaucoup plus élevé que la thèse libertarienne. Celle-ci nous conduit vers des questions délicates sur l'identité psycho-physique, et si l'on veut, à une sorte de « mystère ». Mais l'incompatibilisme dur a un défaut bien plus grand, puisqu'il renonce à nos intuitions sur la responsabilité morale. S'il faut choisir entre le mystérieux et l'inconcevable, je préfère le mystérieux. La science n'aime pas les réalités mystérieuses, car elle voudrait tout expliquer. Mais elle ne nie pas leur existence, quand le mystère lui résiste. Ainsi, nous croyons que les animaux peuvent souffrir, bien que nous ne puissions ni percevoir leurs sensations, ni les réduire à des entités mieux connues. De même, l'acte libre est un mystère, parce qu'il est causalement indéterminé. Mais il pourrait bien être de ces mystères que l'intelligence humaine doit accepter malgré elle.

Bibliographie

- Adams, Robert M. 1994. *Leibniz: Determinist, Theist, Idealist*, Oxford University Press, Oxford.
- Anscombe, G. E. M. 1977. "Soft Determinism", in G. Ryle (ed.), *Contemporary Aspects of Philosophy*, London, 1977. Repris dans *The Collected Philosophical Papers of G. E. M. Anscombe*, vol. 2, Basil Blackwell, Oxford, 1981.
- 1971. "Causality and Determination", leçon inaugurale à l'université de Cambridge, in *The Collected Philosophical Papers of G. E. M. Anscombe*, vol. 2, Basil Blackwell, Oxford, 1981.
 - 1958. *L'intention*, trad. fr. Mathieu Maurice et Cyrille Michon, Gallimard, Paris, 2001.
- Aquin, Thomas d'. *Somme théologique*, trad. fr. aux éditions du Cerf, Paris (1984-1986) : Ia, Q.18, 19, et 86.
- Aristote. *Métaphysique*, trad. fr. Jules Tricot, Vrin, Paris, 1953 : Livre Θ.
- *Éthique à Nicomaque*, trad. fr. Jules Tricot, Vrin, Paris, 1959 : Livre VII.
 - *De l'Âme*, trad. fr. Richard Bodéüs, Flammarion, Paris, 1999.
- Austin, J. L. 1961. "If and Cans", in *Philosophical Papers*, ed. J. O. Urmson and G. Warnock, Clarendon Press, Oxford:153-80.
- Barrow, John. 1991. *Theories of Everything. The Quest for Ultimate Explanation*, Oxford University Press, Oxford. Traduction française : *La grande théorie*, trad. Michel Cassé, Loïc Cohen et Guy Paulus, Albin Michel, Paris, 1994 (puis Flammarion, Paris, 1996).
- Beebe, Helen. 2008. "Smilansky's alleged Refutation of Compatibilism", *Analysis* 68/3: 258-260.
- 2003. "Local Miracle Compatibilism", *Noûs* 37/2: 258-277.
- Beebe, H. and A. Mele. 2002. "Humean Compatibilism", *Mind* 111: 201-223. Repris dans Fischer (2005a).
- Berofsky, Bernard. 2006. "Classical Compatibilism: Not Dead Yet". In Widerker et McKenna (2006).
- 2002. "Ifs, Cans, and Free Will: The Issues". In Kane (2002b): 181-201.
 - 1987. *Freedom from Necessity*, Routledge & Kegan Paul, London.
- Bishop, John. 1983. "Agent-Causation", *Mind* 92: 61-79.
- Bishop, Robert. 2002. "Chaos, Indeterminism, and Free Will", in Kane (2002b): 111-126.
- Blum, Alex. 2003. "The Core of the Consequence Argument", *Dialectica* vol.57 n°4 (2003): 423-429.
- Bok, Hilary. 1998. "Freedom and Practical Reason", extrait de *Freedom and Responsibility*, Princeton University Press, Princeton. Repris dans Watson (2000) et dans Fischer (2005a).
- Boulnois, Olivier. 2002. "Le refoulement de la liberté d'indifférence et les polémiques anti-scotistes de la métaphysique moderne", *Les Études philosophiques* n°2 (2002): 199-237.
- Bratman, M. E. 2007. *Structures of Agency*, Oxford University Press, Oxford.
- 1996. "Identification, Decision, and Trating as a Reason", *Philosophical Topics* 24/2: 1-18. Repris dans Ekstrom (2001).
- Brogie, Louis de. 1937. *La physique nouvelle et les quanta*, Flammarion, Paris.

- Bunge, Mario. 1973. *Philosophy of Physics*, D. Reidel Publishing Company, Dordrecht, Hollande. Trad. française : *Philosophie de la physique*, trad. Françoise Balibar, Seuil, Paris, 1975.
- Byrd, Jeremy. 2007. "Moral Responsibility and Omissions", *The Philosophical Quarterly* 226: 56-67.
- Carlson, Erik. 2003. "Counterexamples to Principle Beta : A Response to Crisp and Warfield", *Philosophy and Phenomenological Research* 66/3: 730-737.
- Chisholm, Roderick. 1995. "Agents, Causes, and Events: The Problem of Free Will", In O'Connor (1995a).
- 1976. *Person and Object: A Metaphysical Study*, Muirhead Library of Philosophy. Open Court, Chicago and La Salle, Illinois.
 - 1964. "Human Freedom and the Self", The Lindley Lecture: 3-15. Repris dans Ekstrom (2001) et dans Watson (2003).
- Clarke, Randolph. 2003. *Libertarian Accounts of Free Will*, Oxford University Press, Oxford.
- 2002. "Libertarian Views: Critical Surveys of Noncausal and Event-Causal Accounts of Free Agency", in Kane (2002b).
 - 1994. "Ability and Responsibility for Omissions", *Philosophical Studies* 73: 195-208.
 - 1993. "Toward a credible Agent-Causal Account of Free Will", *Noûs* 27: 191-203. Repris dans O'Connor (1995a) et dans Watson (2003).
- Cohen, Bernard I. 1960. *The Birth of a New Physics*, W. W. Norton & Company, New York. Traduction française : *Les origines de la physique moderne*, trad. J. Métadier et G. Jeanmougin, éditions du Seuil, Paris, 1985.
- Corcoran, Kevin. 2005. "The Constitution View of Persons", in *In Search of the Soul, Four Views of the Mind-Body Problem*, ed. Joel B. Green and Stuart L. Palmer, InterVarsity Press, Downers Grove, Illinois.
- Cover, Jan, and John O'Leary Hawthorne. 1996. "Free Agency and Materialism", in *Faith, Freedom, and Rationality*, ed. Jeff Jordan and Daniel Howard-Snyder, Rowman & Littlefield Publishers, Inc. Lanham, Maryland.
- Davidson, Donald. 1992. *Paradoxes de l'irrationalité*, trad. Pascal Engel, Éclat, Paris.
- 1980. *Essays on Actions and Events*, Clarendon Press, Oxford. Trad. fr. : *Actions et Événements*, trad. Pascal Engel, PUF, Paris, 1993.
- Dennett, Daniel. 2003. *Freedom Evolves*, Penguin, trad. fr. Christian Cler : *Théorie évolutionniste de la liberté*, Odile Jacob, Paris, 1994.
- 1984. "I could not have done otherwise – So what ?", *The Journal of Philosophy* 81/10: 553-567. Repris dans Kane (2002a).
 - 1984. *Elbow Room, The Varieties of Free Will worth wanting*, Clarendon Press, Oxford.
 - 1978. "On giving Libertarians what they say they want". In Dennett: *Brainstorms*. Bradford Books, Cambridge, Massachussetts. Repris dans O'Connor (1995a).
- Descombes, Vincent. 2002. "Comment savoir ce que je fais ?" *Philosophie* 76: 15 et s.
- Double, Richard. 2002. "Metaethics, Metaphilosophy, and Free Will Subjectivism", in Kane (2002b).
- 1988. "Libertarianism and Rationality". *The Southern Journal of Philosophy* 26: 431-439. Repris dans O'Connor (1995a).
- Dretske, Fred I. 1977. "Laws of Nature", *Philosophy of Science* 44: 248-268.
- Duns Scot, Jean. *Commentaire des Sentences* II, d.25, *Reportata parisiensia* II, d.25, et *Opus Oxoniense* II, d.25. Ces textes sont traduits en français et réunis dans un même

- volume : *Jean Duns Scot, La cause du vouloir*, suivi de *L'objet de la jouissance*, trad., présentation et notes de François Loiret, Les Belles Lettres, Paris, 2009.
- Dupré, John. 1993. *The Disorder of Things, Metaphysical Foundations of the Disunity of Science*, Harvard University Press, Cambridge, Massachussets.
- Eccles, John C. 1994. *How the Self controls its Brain*. Springer-Verlag. Berlin-Heidelberg-New York. Trad. fr. Béatrice Laroche et Jean-Mathieu Luccioni : *Comment la conscience contrôle le cerveau*, Arthème Fayard, Paris, 1997.
- 1991. *Evolution of the Brain : Creation of the Self*, Routledge, London. Trad. fr. : *Évolution du cerveau et création de la conscience*, trad. Jean-Mathieu Luccioni, Flammarion, Paris, 1993.
- Eddington, Arthur. 1928. *The Nature of the Physical World*, McMillan, London.
- Edwards, Jonathan. *Freedom of the Will*, 1754, in *The Works of Jonathan Edwards*, Carlisle, PA, 1834.
- Edwards, Paul. 1958. "Hard and Soft Determinism", in *Determinism and Freedom in the Age of Modern Science*, ed. Sidney Hook, New York University Press, New York. Repris dans Kane (2002a).
- Ekstrom, Laura W. 2002. "Libertarianism and Frankfurt-style Cases", in Kane (2002b).
- 2001. *Agency and Responsibility, Essays on the Metaphysics of Freedom*, ed. Laura Ekstrom, Westview Press, Boulder, Colorado.
 - 2000. "Indeterminist Free Action", in Ekstrom (2001).
 - 2000. *Free Will, A Philosophical Study*, Westview Press, Boulder, Colorado.
- Épicure, *Lettre à Ménécée*, § 133-135.
- *Lettre à Hérodote*, trad. Maurice Solovine, in *Épicure, Doctrines et Maximes*, Hermann & C^{ie}, Paris, 1965: § 39.
- Fischer, John M. 2007. *Four Views on Free Will*. Avec R. Kane, D. Pereboom, et M. Vargas. Blackwell. Malden, Massachussets.
- 2006a. "Responsibility and Alternative Possibilities", in Widerker et McKenna (2006).
 - 2006b. "Responsibility and Agent-Causation", in Widerker et McKenna (2006).
 - 2006c. *My Way: Essays on Moral Responsibility*, Oxford University Press, New York.
 - 2005a. *Free Will: Critical Concepts in Philosophy*, Routledge, New York.
 - 2005b. "Frankfurt Examples: The Moral of the Stories", in David Copp, *Oxford Handbook on Ethical Theory*, Oxford University Press (2005), Oxford. Repris dans Fischer (2005a), vol.3.
 - 2004. "The Transfert of Non-Responsibility", in J. Campbell, M. O'Rourke, and D. Shier (ed.), *Freedom and Determinism: Topics in Contemporary Philosophy Series*, vol.2, MIT Press (2004): 189-209. Repris dans Fischer (2006c).
 - 2004. "Responsibility and Manipulation", *The Journal of Ethics*, vol.8 n°2 (2004): 145-177. Repris dans Fischer (2006c).
 - 2003. "'Ought-Implies-Can", Causal Determinism, and Moral Responsibility", *Analysis*, vol.63 n°3 (juillet 2003): 244-250. Repris dans Fischer (2006c).
 - 2002. "Frankfurt-type Compatibilism", in *The Contours of Agency: Essays on Themes from Harry Frankfurt*, ed. Sarah Buss and Lee Overton, MIT press. Repris dans Watson (2003).
 - 1999. "Frankfurt-style Examples: Responsibility and Semi-Compatibilism", *Ethics* 110. Repris dans Kane (2002a).
 - 1996. "A New Compatibilism", *Philosophical Topics*, vol.24, n°2 (1996). Repris dans Ekstrom (2001).

- 1994. *The Metaphysics of Free Will*, Blackwell, Oxford.
- Fischer, John M., et Mark Ravizza. 1998. *Responsibility and Control*, Cambridge University Press, Cambridge.
- 1993. *Perspectives on Moral Responsibility*, ed. J-M. Fischer and M. Ravizza. Cornell University Press, Ithaca, New York.
- 1992. "When the Will is Free", *Philosophical Perspectives*, ed. J. Tomberlin, vol.6, Ridgview Publishing, Atascadero, CA: 423-451. Repris dans O'Connor (1995a).
- Fischer, John M., et Eleonore Stump. 2000. "Transfer Principles and Moral Responsibility", *Philosophical Perspectives* 14: 47-56.
- Fisette Denis, et Pierre Poirier, 2000. *Philosophie de l'Esprit, État des lieux*, Vrin, Paris.
- Frankfurt, Harry. 1994. "An Alleged Assymetry between Actions and Omissions", *Ethics* 104: 620-623.
- 1992. "The Faintest Passion", Presidential Address, *Proceedings of the American Philosophical Association* 66: 5-16. Repris dans Fischer (2005a), vol.4.
- 1987. "Identification and Wholeheartedness", in Schoeman, *Responsibility, Character and Emotion*. Cambridge University Press, Cambridge, 1987. Repris dans Fischer et Ravizza (1993) et dans Fischer (2005a), vol. 4.
- 1971. "Freedom of the Will and the Concept of a Person", *Journal of Philosophy* 68: 5-20. Repris dans Ekstrom (2001), dans Kane (2002a), et dans Fischer (2005a), vol.4.
- 1969. "Alternate Possibilities and Moral Responsibility", *Journal of Philosophy* 66: 829-839. Repris dans Watson (2003).
- Garson, James. 1993. « Chaos and Free Will », conférence donnée à l'American Philosophical Association, Pacific Division meeting, march 1993, et publiée dans *Philosophical Psychology*, vol. 8, issue 4 (1995): 365-374.
- Gilson, Étienne. 1913. *La Liberté chez Descartes et la Théologie*, F. Alcan, Paris. Rééd. Vrin (1983).
- Ginet, Carl. 2002. "Reasons Eplanations of Action: Causalist versus Noncausalist Accounts", in Kane (2002b).
- 2000. "The Epistemic Requirements for Moral Responsibility", *Philosophical Perspectives* 14: 267-277.
- 1997. "Freedom, Responsibility and Agency", *The Journal of Ethics* 1/1 (1997): 374-380. Repris dans Kane (2002a).
- 1996. "In Defense of the Principle of Alternative Possibilities: Why I Don't find Frankfurt's Argument Convincing", *Philosophical Perspectives* 10: 403-417. Repris dans Widerker et McKenna (2006), et dans Fischer (2005a), vol.3.
- 1990. *On Action*, Cambridge University Press, Cambridge.
- 1989. "Reasons Explanation of Action: An Incompatibilist Account", *Philosophical Perspectives* 3. Repris dans O'Connor (1995a).
- 1966. "Might we have no Choice ?", in Lehrer, *Freedom and Determinism*, Random House, New York, 1966. Repris dans Fischer (2005a), vol.2.
- Glauser, Rischard. 2003. "Thinking and Willing in Locke's Theory of Human Freedom", *Dialogue* 42: 695-724.
- Gnassounou, Bruno. 2006. "Possibilités conditionnelles", in *Les Dispositions en Philosophie et en Science*, ed. Bruno Gnassounou et Max Kistler, éditions du CNRS, Paris.
- 2005. "Pouvoir et possibilité", in *Causes, Pouvoirs, Dispositions en Philosophie*, ed. Bruno Gnassounou et Max Kistler, PUF, Paris.
- Haji, Ishtiyaque. 2002. "Compatibilist View of Freedom and Responsibility", in Kane (2002b).

- 1992. "A Riddle regarding Omissions", *Canadian Journal of Philosophy* 22: 485-502.
- Haji, Ishtiyaque & Michael McKenna. 2004. "Dialectical Delicacies about Freedom and Alternative Possibilities", *Journal of Philosophy* 101/6: 299-314. Repris dans Fischer (2005a).
- Hasker, William. 1999. *The Emergent Self*, Cornell University Press, Ithaca.
- Hempel, Carl. 1948. "Studies in the Logic of Explanation", *Philosophy of Science* 15: 135-175. Repris dans Carl Hempel, *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, The Free Press, New York, 1965.
- Hippone, Augustin d'. *Du Libre arbitre*, in St Augustin, *Oeuvres III : Philosophie, catéchèse, polémique*, La Pléiade, Gallimard, Paris, 2002.
- Hobbes, Thomas. *De la liberté et de la nécessité*. Introduction, traduction, notes, index et glossaires par Franck Lessay, Vrin, Paris, 1993. Première édition en anglais en 1656 à Londres.
- Hodgson, David. 2002. "Quantum Physics, Consciousness and Free Will", in Kane (2002b).
- 1999. "Chess, Life and Superlife", *The Journal of Consciousness Studies*, vol.6 n°8-9. Repris dans Kane (2002a).
- Honderich, Ted. 2002. "Determinism as True, Compatibilism and Incompatibilism as False, and the Real Problem", in Kane (2002b).
- 1993. *How free are you ?* Oxford University Press, Oxford.
- 1988. *A Theory of Determinism*, 2 vols, Clarendon Press, Oxford.
- 1973. *Essays on Freedom and Action*. ed. Ted Honderich. Routledge & Kegan Paul. London.
- Hume, David. *Enquête sur l'Entendement humain*, trad. fr. Philippe Baranger et Philippe Saltel, Aubier-Montaigne, Paris, 1947. Rééd. Flammarion, Paris, 1987.
- Hunt, David. 2005. "Moral Responsibility and buffered alternatives", *Midwest Studies in Philosophy*. Repris dans Fischer (2005a), vol.3.
- 2000. "Moral Responsibility and Unavoidable Action", *Philosophical Studies* 97: 195-227. Repris dans Fischer (2005a) vol.3.
- James, William. 1888. "What the Will effects", *Scribner's Magazine* 1/3: 355-366.
- 1884. "The Dilemma of Determinism", *Unitarian Review* (sept. 1884).
- Kane, Robert. 2003. "Responsibility, Indeterminism and Frankfurt-style Cases: A Reply to Mele and Robb", in Widerker et McKenna (2006).
- 2002a. *Free Will*, ed. R. Kane. Blackwell, Malden, Massachussets.
- 2002b. *The Oxford Handbook of Free Will*, ed. R. Kane, Oxford University Press, Oxford.
- 2002c. "Free Will. New Directions for an Ancient Problem", in Kane (2002a). Repris dans Fischer, *Free Will: Critical Concepts in Philosophy* (2005a) vol.3.
- 2002d. "Some Neglected Pathways in the Free Will Labyrinth", in Kane (2002b).
- 1999. "Responsibility, Luck and Chance: Reflections on Free Will and Indeterminism", *Journal of Philosophy* 97/6: 217-240. Repris dans Ekstrom (2001) Watson (2003), et Fischer (2005a) vol.3.
- 1996. *The Significance of Free Will*, Oxford University Press, Oxford.
- 1989. "Two Kinds of Incompatibilism", *Philosophy and Phenomenological Research* 50: 219-254. Repris dans O'Connor (1995a).
- Kant, Emmanuel. *Critique de la Raison Pure*, trad. fr. A. Tremesaygues et B. Pacaud, PUF, Paris, 1965. Première édition allemande en 1784.
- Kapitan, Tomis. 2002. "A Master Argument for Incompatibilism ?", in Kane (2002b).
- Kenny, Anthony. 1975. *Will, Freedom and Power*, Blackwell, Oxford.

- Kim, Jaegwon. 2005. *Trois Essais sur l'Émergence*, trad. Mathieu Mulcey, Ithaque, Paris.
- Kistler, Max. 1999. *Causalité et Lois de la Nature*, Vrin, Paris.
- Lehrer, Keith. 1980. "Preferences, Conditionals and Freedom", in Peter Van Inwagen (ed), *Time and Cause, Essays presented to Richard Taylor, Philosophical series studies in philosophy* vol.19 (1980): 76-96. Repris dans Fischer (2005a) vol.2.
- 1976. "'Can' in theory and practice: a possible worlds analysis, in *Action Theory: Proceedings of the Winnipeg Conference on Human Action*, ed. M. Brand et D. Walton, Reidel, Dordrecht: 241-270. Repris dans Fischer, *Free Will: Critical Concepts in Philosophy* (2005a) vol.2.
 - 1968. "'Can's' Without 'If's'", *Analysis* 29: 29-32.
 - 1966. "Freedom and the Power of Preference". In *The Determinism and Freedom Website*. ed. Ted Honderich.
 - *Freedom and Determinism*, Keith Lehrer (ed.), New York, 1966, p. 196.
- Leibniz, G. W. *Discours de Métaphysique*, in *Leibniz, Discours de Métaphysique et autres textes*, ed. Christiane Frémont, Flammarion, Paris, 2001.
- *Conversation avec Sténon*, in *Leibniz, Discours de Métaphysique et autres textes*, ed. Christiane Frémont, Flammarion. Paris, 2001.
 - *Conversation sur la liberté et le destin*, in *Leibniz, Principes de la Nature et de la Grâce, Monadologie et autre textes*, ed. Christiane Frémont, Flammarion, Paris (1996).
 - *Correspondance avec Arnauld*, in *Leibniz, Discours de Métaphysique et Correspondance avec Arnauld*, ed. Jean-Baptiste Rauzy, Pocket, Paris, 1993.
 - *Nouveaux Essais sur l'Entendement humain*, ed. Jacques Brunschwig, Flammarion, Paris, 1990.
- Lewis, David. 1981. "Are We Free to Break the Laws ?" *Theoria* 47: 113-121. Repris dans Watson (2003) et dans Fischer, *Free Will, Critical Concepts in Philosophy* (2005a) vol.2.
- Libet, Benjamin. 2002. "Do We Have Free Will ?" in Kane (2002b).
- Locke, John. *Essai sur l'Entendement humain*, trad. Jean-Michel Vienne, Vrin, Paris, 2001-2003.
- Lucrèce, *De Natura Rerum*, trad. A. Ernout, Belles Lettres, Paris, 1924.
- McIntyre, Alison. 1994. "Compatibilists Could Have Done Otherwise: Responsibility and Negative Agency", *The Philosophical Review*, vol.103/3: 453-488.
- McKay, Thomas, et David Johnson. 1996. "A Reconsideration of an Argument against Compatibilism", *Philosophical Topics* 24: 113-122.
- McKenna, Michael. 2003. "Robustness, Control, and the Demand for Morally Significant Alternatives", in Widerker et McKenna (2006). Repris dans Fischer, *Free Will: Critical Concepts in Philosophy* (2005a) vol.3.
- Mele, Alfred. 2003. "Soft Libertarianism and Flickers of Freedom", in Widerker et McKenna (2006): 251-264.
- 2002. "Autonomy, Self-Control, and Weakness of Will", in Kane (2002b).
 - 1995. *Autonomous Agents: From Self-Control to Autonomy*, Oxford University Press, New York.
- Mele, Alfred, et David Robb. 2003. "Bbs, magnets, and seesaws: the metaphysics of Frankfurt-style Cases", in Widerker et McKenna (2006). Repris dans Fischer, *Free Will: Critical Concepts in Philosophy* (2005a), vol.3.
- 1998. "Rescuing Frankfurt-style Cases". *Philosophical Review* 107: 97-112. Repris dans Ekström (2001) et dans Fischer (2005a), vol.3.
- Michon, Cyrille. 2004. *Prescience et Liberté*, PUF, Paris.

- Moore, G. E. 1912. "Free Will". À l'origine, ch. VI de *Ethics*, William & Norgate, London, 1912. Trad. fr. par Michel Gouverneur dans Moore, *Principia Ethica*, PUF, Paris, 1998: 329-343.
- Murray Michael. "Pure Omissions, Responsibility, and Character". Inédit.
- Murray, Michael, et Sam Ruchlewicz. "Without a Trace". Inédit.
- Nagel, Thomas. 1986. *The View from Nowhere*, Oxford University Press, New York. Deux extraits de ce livre sont repris dans O'Connor (1995a) et Watson (2003).
- Nielsen, Kai. 1971. "The Compatibility of Freedom and Determinism", in Nielsen, *Reason and Practice*, Harper & Row, New York. Repris dans Kane (2002b).
- Nozick, Robert. 1981. "Choice and Indeterminism", extrait de R. Nozick, *Philosophical Explanations*, Harvard University Press, Cambridge, Massachussets, 1981. Repris dans O'Connor (1995a).
- O'Connor, Timothy. 2003. "Groundworks for an Emergentist Account of the Mental". *PCID* 2.3.1.
- 2002. "Libertarian Views: Dualist and Agent-Causal Theories", in Kane (2002b).
 - 2000. *Persons and Causes: The Metaphysics of Free Will*, Oxford University Press, Oxford.
 - 1998. "The Agent as Cause". In *Metaphysics: The Big Questions*, ed. Peter Van Inwagen et Dean Zimmerman, Blackwell, Oxford. Repris dans Kane (2002a).
 - 1995a. *Agents, Causes and Events: Essays on Free Will and Indeterminism*, ed. Timothy O'Connor, Oxford University Press, Oxford.
 - 1995b. "Agent Causation", in O'Connor (1995a). Repris dans Watson (2003).
 - 1993. "On the Transfert of Necessity", *Nous* 27: 204-218.
- Otsuka, Michael. 1998. "Incompatibilism and the Avoidability of Blame", *Ethics* 108: 685-701.
- Parfit, Derek. 1984. *Reasons and Persons*, Oxford University Press, Oxford.
- Pereboom, Derk. 2002. "Meaning in Life without Free Will", *Philosophic Exchange* 33 (2002-3): 18-34.
- 2001. *Living without Free Will*, Cambridge University Press, Cambridge.
 - 2000. "The Explanatory Irrelevance of Alternative Possibilities", in Kane (2002a), extrait de "Alternatives Possibilities and Causal Histories", *Philosophical Perspectives* 14.
- Pettit, Philip, and Michael Smith. 1996. "Freedom in Belief and Desire", *Journal of Philosophy* 93/9: 429-449. Repris dans Watson (2003).
- Plantinga, Alvin. 1974. *The Nature of Necessity*, Oxford University Press, Oxford.
- Platon, *Phèdre*, trad. fr. Émile Chambry, Flammarion, Paris, 1993.
- Popper, Karl. 1990. *A World of Propensities*, Thoemmes Press, Bristol, trad. fr. Alain Boyer: *Un univers de propensions*, L'Éclat, Combas, 1992.
- Prigogine, Ilya. 1990. "Loi, histoire... et désertion", in Pomian (1990a).
- Prigogine, Ilya, et Isabelle Stengers. 1990. "La querelle du déterminisme, six ans après", in Pomian (1990a).
- Ravizza, Mark. 1994. "Semi-Compatibilism and the Transfert of Nonresponsibility", *Philosophical Studies* 75.
- Reid, Thomas. *Essays on The Active Powers of the Human Minds: An Inquiry into the Human Mind on the Principle of Common Sense*, R. Griffin & Co, Glasgow, 1843. Première édition en anglais en 1788.
- Renouvier, Charles. *Les dilemmes de la métaphysique pure*, F. Alcan, Paris, 1900: ch. IV: "Le déterminisme et le libre arbitre".

- Rowe, William. 2003. "Alternate Possibilities and Reid's Theory of Agent-Causation", in Widerker et McKenna (2006): 219-234.
- 1991. "Responsibility, Agent-Causation, and Freedom: An Eighteen Century View", *Ethics* 101: 237-257. Repris dans Fischer et Ravizza (1993).
 - 1989. "Causing and Being Responsible for What Is Inevitable", *American Philosophical Quarterly* 26: 153-159. Repris dans Fischer et Ravizza (1993).
 - 1987. "Two Concepts of Freedom", *The Proceedings and Adresses of the American Philosophical Association* 61: 43-64. Repris dans O'Connor (1995a).
- Sartorio, Carolina. 2009. "Omissions and Causalism", *Noûs* 43/3 (2009): 513-530.
- 2005. "A New Asymmetry Between Actions and Omissions", *Noûs* 39/3 (2005): 460-482.
- Saunders, John T. 1968. "The Temptations of "Powerlessness"", *American Philosophical Quarterly* 5 (1968): 219-225. Repris dans Fischer (2005a), vol.2.
- Scanlon, T. M. 1988. "The Significance of Choice", in *The Tanner Lectures on Human Values*, ed. Sterling M. McMurrin (1988): 151-177.
- Searle, John R. 2007. *Freedom and Neurobiology*, Columbia University Press. New York.
- 1984. *Minds, Brains, and Science*, Harvard University Press, Cambridge, Massachussetts.
- Sher, George. 2006. "Out of Control", *Ethics* 116: 285–301.
- Shoemaker, Sydney, and Richard Swinburne. 1984. *Personal Identity*, Blackwell, Oxford.
- Skinner, B. F. 1977. "Walden Two: Freedom and the Behavioral Sciences", extrait de *Walden* 2. Pearson Educational, Inc. Upper Saddle River, New Jersey, 1977. Repris dans Kane (2002b).
- Slote, Michael. 1982. "Selective Necessity and the Free Will Problem", *Journal of Philosophy* 79: 5-24. Repris dans Fischer (2005a), vol.2: 111-125.
- Smart, J. J. C. 1961. "Free Will, Praise and Blame", *Mind* 70: 291-306. Repris dans Watson (2003).
- Smilansky, Saul. 2007. "Determinism and prepunishment: the radical nature of compatibilism", *Analysis* 67/4 (2007): 347-349.
- 2002. "Free Will, Fundamental Dualism, and the Centrality of Illusion." In Kane (2002b).
- Smith, Patricia G. 1990. "Contemplating Failure: The Importance of Unconscious Omission", *Philosophical Studies*, 59/2 (1990): 159-176.
- Sorabji, Richard. 1980. *Necessity, Cause and Blame: Perspectives on Aristotle's Theory*, The University of Chicago Press, Chicago.
- Strawson, Galen. 2002. "The Bounds of Freedom", in Kane (2002b).
- 1994. "The Impossibility of Moral Responsibility", *Philosophical Studies* 75/1-2: 5-24. Repris dans Watson (2003).
 - 1986. "Libertarianism, Action and Self-Determination", extrait de *Freedom and Belief*, Clarendon Press, Oxford, 1986. Repris dans O'Connor (1995a).
 - 1986. "On "Freedom and Resentment"", extrait de *Freedom and Belief*, Clarendon Press, Oxford, 1986. Repris dans Fischer et Ravizza (1993).
- Strawson, Peter. 1962. "Freedom and Ressentment", *Proceedings of the British Academy* 48: 1-25. Repris dans Fischer et Ravizza (1993), dans Ekstrom (2001), et dans Watson (2003).
- Stump, Eleonore. 2003. "Moral Responsibility without Alternative Possibilities", in Widerker et McKenna (2006).

- 2001. "Augustine on Free Will", in Eleonore Stump et Norman Kretzman (ed.), *Cambridge Companion to Augustine*, Cambridge University Press, Cambridge, 2001: 124-147.
- 1999. "Alternative Possibilities and Moral Responsibility : the Flicker of Freedom", *The Journal of Ethics* 3: 299-324.
- 1995. "Non-Cartesian Dualism and Materialism without Reductionism", *Faith and Philosophy* 12 (1995): 505-531.
- 1993. "Sanctification, Hardening of the Heart, and Frankfurt's Concept of Free Will", in Fischer et Ravizza (1993).
- 1990. "Intellect, Will, and the Principle of Alternate Possibilities", in *Christian Theism and the Problems of Philosophy*, ed. Michael D. Betty, University of Notre Dame Press. Repris dans Fischer et Ravizza (1993).
- Suarez, Francisco. *Disputationes metaphysicae*, Disputatio XVIII, section X (au vol. XXV des *Opera omnia*, ed. Louis Vivès, Paris, 1861).
- Swinburne, Richard. 1997. *The Evolution of the Soul*, Clarendon Press, Oxford. (édition revue et corrigée).
- Taylor, Christopher, and Daniel Dennett. 2002. "Who's Afraid of Determinism ? Rethinking Causes and Possibilities", in Kane (2002b).
- Taylor, Richard. 1976. "Responsibility for Self", in A. E. Rorty (ed.), *The Identities of Persons*, University of California Press, Berkeley. 1976.
- 1958. "Determination and the Theory of Agency", in Sidney Hook ed, *Determinism and Freedom in the Age of Modern Physics*, New York University Press, New York.
- Timpe, Kevin. 2007. "Source Incompatibilism and its Alternatives", *American Philosophical Quarterly* 44/2 (2007): 143-145.
- 2006. "A Critique of Frankfurt-Libertarianism", *Philosophia* 34: 189-202.
- Van Fraassen, Bas. *Lois et Symétrie*, présentation et trad. fr. par Catherine Chevalley, Vrin, Paris, 1994. Première édition en anglais : *Laws and Symmetry*, Oxford University Press, Oxford, 1989.
- Van Inwagen, Peter. 2002. "Free Will remains a Mystery", in Kane (2002b).
- 1998. "The Mystery of Metaphysical Freedom", in *Metaphysics: The Big Questions*. ed. Peter Van Inwagen and Dean Zimmerman. Repris dans Kane (2002a).
- 1994. "When Is the Will Not Free", *Philosophical Studies* 75: 95-114.
- 1990. *Material Beings*, Cornell University Press, Ithaca, New York.
- 1989. "When Is the Will Free ?" in *Philosophical Perspectives*, vol.3. ed. J. Tomberlin. Ridgeview. Atascadero, CA, 1989: 399-422. Repris dans O'Connor (1995a).
- 1983. *An Essay on Free Will*, Clarendon Press, Oxford.
- 1975. "The Incompatibility of Free Will and Determinism", *Philosophical Studies* 27: 185-199. Repris dans Ekstrom (2001) et Kane (2002a).
- Velleman, J. David. 1992. "What Happens When Someone Acts ?", *Mind* 101: 461-481. Repris dans Fischer et Ravizza (1993).
- Wallace, R. Jay. 1999. "Addiction as Defect of the Will: Some Philosophical Reflections", *Law and Philosophy* 18/6: 621-654. Repris dans Watson (2003).
- Walter, Henrik. "Neurophilosophy of Free Will", in Kane (2002b).
- Watson, Gary. 2003. *Free Will*, ed. Gary Watson, Oxford University Press, Oxford.
- 1987. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme", in *Responsibility, Character and the Emotions*, ed. Ferdinand David Schoeman, Cambridge University Press, Cambridge, 1987: 256-286. Repris dans Fischer et Ravizza (1993) et dans Kane (2002a).

- 1975. "Free Agency", *Journal of Philosophy* 72/8 (1975): 205-220. Repris dans Ekstrom (2001) et dans Watson (2003).
- Widerker, David. 2009. "Frankfurt-type Incompatibilism", Colloque sur la responsabilité et le libre arbitre à l'université de Nantes, le 14 octobre 2009. Inédit.
- 2006. "Libertarianism and the Philosophical Significance of Frankfurt Scenarios", *Journal of Philosophy* 13/4 (2006): 163-187.
 - 2003. "Blameworthiness and Frankfurt's Argument Against the Principle of Alternative Possibilities", in Widerker et McKenna (2006).
 - 2002. "Responsibility and Frankfurt-type Examples", in Kane (2002b).
 - 2000. "Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look", *Philosophical Perspectives* 14 (2000): 181-201. Repris et corrigé dans « Blameworthiness and Frankfurt's Argument Against the Principle of Alternative Possibilities », in Widerker & McKenna (2006): 53-73.
 - 1995. "Libertarianism and the Frankfurt's Attack on the Principle of Alternative Possibilities", *Philosophical Review* 104: 247-261. Repris dans Ekstrom (2001) et dans Watson (2003).
- Widerker, David, and Michael McKenna (ed.). 2006. *Moral Responsibility and Alternative Possibilities*, Ashgate, Hants, England.
- Wiggins, David. 1973. "Towards a Reasonable Libertarianism", in Honderich (1973): 31-61. Repris dans Watson (2003).
- Wolf, Susan. 1990. *Freedom within Reason*, Oxford University Press, New York.
- 1987. "Sanity and the Metaphysics of Responsibility", extrait de *Responsibility, Character and the Emotions*, Cambridge University Press, Cambridge, 1987: 46-62. Repris dans Watson (2003).
 - 1981. "The Importance of Free Will", *Mind* 90: 386-405. Repris dans Fischer et Ravizza (1993).
- Wyma, Keith. 1997. "Moral Responsibility and Leeway for Action", *American Philosophical Quarterly* 34: 57-70. Repris dans Fischer (2005a), vol.3.
- Zimmerman, David. 1994. "Acts, Omissions, and Semi-Compatibilism", *Philosophical Studies* 73: 209-223.

Table des matières

Sommaire	3
Introduction	5
I La liberté et la contingence	13
1.1 La liberté et le libre arbitre	13
1.2 La contingence et les possibilités alternatives.....	17
1.3 Les différentes conceptions du libre arbitre	21
1.4 Différents sens du mot « pouvoir »	24
1.5 Le pouvoir de faire l'impossible	27
1.6 Trois espèces de capacités.....	30
1.7 La liberté et les dispositions.....	35
1.8 « Causer » et « rendre nécessaire »	39
1.9 Causes et défauts	44
1.10 La nécessité compatible avec le libre arbitre ?.....	46
1.11 Le restrictivisme et la liberté d'indifférence.....	49
1.12 Résumé.....	55
II La responsabilité et le pouvoir des alternatives	59
2.1 Les exemples à la Frankfurt	61
2.2 Le principe de prévention possible (PPP)	63
2.3 Le frémissment de liberté : un résidu des possibilités alternatives.....	68
2.4 L'argument du dilemme.....	75
2.5 Un exemple sans intervenant contrefactuel.....	79
2.6 Surdétermination et blocage parfait	84
2.7 Valeur et portée du vacillement de liberté. La « robustesse » des possibilités alternatives ..	95
2.8 Résumé.....	103

III L'argument du transfert 107

3.1 Introduction. L'argument du transfert (ou de la « conséquence »).....	107
3.2 La fixité des lois de la nature (FL)	109
3.3 La fixité du passé (FP)	118
3.4 L'argument du transfert est-il valide ?.....	123
3.5 L'argument du transfert prouve-t-il l'incompatibilité du déterminisme et de la responsabilité morale ?.....	131
3.6 La stratégie diodorienne.....	136
3.7 Le transfert de non-responsabilité (TNR)	146
3.8 Les cas de causalité simultanée.....	150
3.9 Première condition de la responsabilité : la contribution de l'agent	158
3.10 La contribution volontaire de l'agent. Conditions épistémiques de la responsabilité morale	164
3.11 Résumé.....	171

IV L'analyse conditionnelle des pouvoirs et les théories du « moi profond » 175

4.1 Introduction.....	175
4.2 La définition conditionnelle de la liberté	177
4.3 La définition conditionnelle redoublée. Les niveaux de désirs	182
4.4 Y a-t-il une infinité de niveaux dans les inclinations ?	189
4.5 Application au compatibilisme	195
4.6 Contrôle de soi et identité avec soi.....	198
4.7 Liberté et rationalité	205
4.8 Capacités d'agir mieux et d'agir moins bien. Sont-elles asymétriques ?	209
4.9 Résumé.....	212

V Les omissions et le pouvoir des alternatives..... 215

5.1 Introduction.....	215
5.2 Une théorie libertarienne de la responsabilité dans les omissions	216
5.3 Première difficulté. La distinction entre les omissions et les actions.....	221
5.4 Deuxième difficulté : l'omission peut-elle causer un effet ?	225
5.5 Troisième difficulté. Le débat sur la transitivité de la causalité.....	230
5.6 L'approche réductionniste	237
5.7 Le problème des asymétries et l'argument de l'exclusion	243
5.8 L'analyse conditionnelle de la responsabilité dans les omissions	251
5.9 Un conflit d'intuitions touchant l'asymétrie entre les actions et les omissions.....	262
5.10 Résumé.....	267

VI La sensibilité aux raisons et les mondes possibles.....	271
6.1 Introduction.....	271
6.2 La sensibilité aux raisons	273
6.3 Première objection : un mécanisme ou des raisons implantés artificiellement	276
6.4 Les mondes possibles pertinents. L'impossibilité de trouver un critère	280
6.5 L'objection inverse : un agent sensible aux raisons, mais qui ne se contrôle pas dans le monde actuel	292
6.6 Défauts ou obstacles dans le monde actuel ou dans les mondes possibles	297
6.7 Cinquième objection : si la sensibilité aux raisons est une disposition, on doit pouvoir l'attribuer à l'agent sans se référer à des mondes possibles proches du monde actuel	302
6.8 Causes externes ou internes, causes naturelles ou non-naturelles.....	310
6.9 Causes et raisons. Le déterminisme des raisons.....	312
6.10 Conclusion. La balance des intuitions.....	316
VII Le libre arbitre et le monde physique Déterminisme et indéterminisme	323
7.1 Introduction.....	323
7.2 « Tout a une cause », « rien ne vient du néant ». Le principe de causalité	324
7.3 Toute relation causale est-elle nomologique ?	329
7.4 Le déterminisme peut-il avoir un fondement empirique ?	336
7.5 Le hasard et l'indéterminisme.....	343
7.6 Le libre arbitre et l'indéterminisme quantique. Première difficulté	345
7.7 Le <i>Mind argument</i> : la notion de hasard	348
7.8 Le <i>Mind argument</i> : le contrôle de soi.....	353
7.9 Difficultés pour la théorie de la causalité événementielle.....	361
7.10 La causalité de l'agent	364
7.11 L'indéterminisme s'oppose-t-il au contrôle de soi ?	371
7.12 Le libre arbitre et le physicalisme	377
Conclusion.....	387
Bibliographie.....	393
Table des matières	403

Le déterminisme et la responsabilité morale

Le déterminisme est-il compatible avec la responsabilité morale ? Oui, répondent les compatibilistes. Or cette idée pose une série de problèmes. Pour la soutenir, il faut réfuter deux principes très intuitifs : celui des possibilités alternatives, selon lequel on est responsable de ce qu'on fait seulement si on a eu le pouvoir de ne pas le faire ; et celui du transfert de non-responsabilité (« nul n'est responsable de Y si X cause Y de manière déterministe, et si nul n'est responsable de X »). Même s'il y avait des contre-exemples à ces principes, la responsabilité par omission serait difficilement compréhensible sans l'idée d'un pouvoir des alternatives. Le compatibilisme doit donc accorder une place à l'idée d'un pouvoir des alternatives. Pour le faire sans se contredire, il analyse ce pouvoir soit de manière conditionnelle (S peut faire A signifierait « si S voulait faire A, il ferait A »), soit en termes de mondes possibles. La première analyse est un cas particulier de la deuxième, car elle revient à dire que S peut faire A s'il y a des mondes possibles où S fait A volontairement. Pour qu'un critère de ce genre ne soit pas trop souple, il faut préciser que seuls des mondes possibles proches du monde actuel sont pertinents du point de vue de la responsabilité morale. Mais comment délimiter le domaine des mondes possibles pertinents ? D'un point de vue compatibiliste, il semble impossible de résoudre ce problème de manière satisfaisante. Finalement, cette thèse ne paraît plausible que si le libertarisme est encore moins intuitif. La principale question, pour le libertarisme, sera de savoir comment l'indéterminisme physique (à l'intérieur du cerveau) permet la responsabilité morale au lieu de l'exclure.

Mots-clés : libre arbitre, compatibilisme, responsabilité morale, déterminisme, possibilités alternatives, indéterminisme, causalité de l'agent, principe du transfert.

Determinism and moral responsibility

Is determinism compatible with moral responsibility ? Compatibilism answers "yes". But such a thesis entails a series of problems. It requires the falsity of two very intuitive principles : the principle of alternative possibilities, i.e. the assertion that one is responsible for what they do only if they have the power not to do it ; and the principle of the transfer of non-responsibility ("if no one is responsible for X and if X causes Y deterministically, no one is responsible for Y"). Assuming that there are counterexamples to these principles, responsibility by omission is hardly understandable without the idea of a power of alternatives. Consequently, compatibilism must attach importance to the power of alternatives. In order to do it without contradiction, compatibilism must define this power either on a conditional mode ("S can do A" would mean "S does A if S want to do A"), or in terms of possible worlds. The first analysis is a particular case of the latter, since it amounts to say that S can do A if there are possible worlds where S do A voluntarily. To prevent such a criterion from being too flexible, one must specify that the possible words pertinent to moral responsibility must be sufficiently similar to the actual world. But how to define the domain of the pertinent possible worlds ? It seems impossible to solve this problem satisfactorily from a compatibilist point of view. In fact, this thesis seems plausible only if libertarianism is even less intuitive. The main question, for libertarianism, will be how physical indeterminism (inside the brain) ensures moral responsibility instead of undermining it.

Keywords : free will, compatibilism, moral responsibility, determinism, alternative possibilities, indeterminism, agent causation, transfer principle.