

## Tables des matières

Résumé .....	iii
Tables des matières.....	iv
Remerciements.....	vi
Introduction .....	1
Chapitre 1 : Qu'est-ce que le libre arbitre ?.....	6
1.0 Introduction.....	6
1.1 Définition générale .....	7
1.1.1 Distinction entre libre arbitre et liberté.....	7
1.1.2 Distinction entre libre arbitre et responsabilité morale.....	8
1.1.3 Libre arbitre et causalité.....	9
1.1.4 Les notions du sens commun : présentation et critique .....	11
1.2 Les notions libertarianistes de libre arbitre .....	15
1.2.1 Le déterminisme.....	16
1.2.2 Le libertarianisme de Laura Ekstrom.....	18
1.3 Les notions compatibilistes de libre arbitre.....	25
1.3.1 Le compatibilisme d'Harry Frankfurt.....	25
1.3.2 Le compatibilisme de John Martin Fischer et Mark Ravizza .....	30
1.4 Conclusion .....	37
Chapitre 2 : Contre le compatibilisme.....	39
2.0 Introduction.....	39
2.1 L'argument de la conséquence .....	40
2.1.1 Le contre-exemple de McKay et Johnson .....	45
2.1.2 Sauver la règle ( $\beta$ ) en précisant la signification de l'opérateur $Np$ .....	46
2.1.3 Un conflit d'intuitions .....	47
2.2 L'argument de la manipulation .....	48
2.2.1 Deux types de modèles de libre arbitre.....	54
2.2.2 Les modèles non historiques de libre arbitre : présentation et critique.....	55
2.2.3 Le modèle historique de Fischer et Ravizza: présentation.....	57

2.2.4 Le modèle historique de Fischer et Ravizza : critique .....	61
2.2.5 Une vue d'ensemble .....	64
2.3 Conclusion .....	65
Chapitre 3 : Contre le libertarianisme .....	67
3.0 Introduction.....	67
3.1 L'argument de la chance .....	67
3.1.1 L'objection d'Ekstrom à l'argument de la chance .....	70
3.2 Est-il possible de structurer son caractère ? .....	72
3.2.1 L'argument de base .....	74
3.2.2 La faculté évaluative échappe-t-elle au déterminisme ? .....	75
3.3 Le quasi-déterminisme .....	77
3.3.1 L'appui des sciences physiques au libertarianisme .....	79
3.3.2 Un manque de preuves empiriques .....	81
3.4 Conclusion .....	83
Chapitre 4 : Pour le scepticisme.....	84
4.0 Introduction.....	84
4.1 Un modèle normatif.....	85
4.1.1 La communauté de responsabilité .....	86
4.2 La critique de Smilansky .....	87
4.2.1 L'hypothèse du monisme .....	90
4.2.2 Un dualisme fondamental .....	91
4.3 La solution : l'illusionnisme .....	93
4.4 Sauver la responsabilité morale .....	96
4.5 Conclusion .....	97
Conclusion .....	99
Bibliographie .....	103

## Remerciements

J'aimerais d'abord remercier ma directrice Renée Bilodeau pour son support et pour la rigueur avec laquelle elle a révisé mes chapitres, et ce malgré les indénombrables lectures que je lui ai infligées. C'est elle qui m'a introduit à la philosophie analytique et qui m'en a transmis la passion. Sans elle, j'aurais probablement abandonné mes études en philosophie depuis longtemps. Renée m'a servi de modèle tout au long de mes études en philosophie et je la remercie chaleureusement pour tout ce qu'elle m'a apporté.

Je veux remercier mes collègues de bureau, Jérôme Brousseau, Kate Blais et Jeffrey Elawani, pour les nombreuses conversations que j'ai eues avec eux à propos de mon mémoire. Merci aussi d'avoir enduré ma musique et d'avoir rendu mon espace de rédaction aussi stimulant.

Je tiens aussi à remercier tous mes amis qui ont pris le temps de lire, de commenter ou de discuter de mon mémoire : Mélyssa Thibodeau-Doré, François-Xavier Ouellet, Olivier Provencher St-Pierre, Sophie Savard-Laroche, David Rocheleau-Houle, Pier-Yves Larouche Maltais et Patrick Fournier.

Je remercie également ma famille pour leur support et leurs encouragements. Je remercie particulièrement mon frère, Charles Ouellet, qui m'a accompagné (littéralement) tout au long de mes études en philosophie.

Enfin, je remercie le GRIN (Groupe de Recherche Interuniversitaire sur la Normativité) et la faculté de philosophie de l'Université Laval pour leur soutien financier.

## Introduction

Dans la vie de tous les jours, nous enchaînons les décisions sans véritable effort. Au réveil, certains décideront délibérément d'ignorer la sonnerie de leur téléphone pour dormir quelques minutes de plus. Les plus vaillants d'entre nous se lèveront immédiatement et iront choisir les morceaux de vêtement qu'ils porteront pour la journée. En examinant les bananes presque trop mûres qui traînent sur le comptoir, certains décideront de les manger pour déjeuner, et d'autres décideront plutôt de cuisiner un pain aux bananes en soirée. La vie regorge de ces moments où les actions se succèdent naturellement. Il arrive, en de plus rares occasions, des événements importants qui valent la peine de s'arrêter et de délibérer plus longuement. Vous vous faites offrir le poste de vos rêves à l'étranger, mais l'offre s'accompagne d'importants sacrifices au niveau personnel. Vous avez l'occasion de vous acheter un condo à très bon prix au centre-ville, mais cela exige que vous laissiez tomber le voyage autour du monde dont vous rêvez depuis toujours. Dans de telles situations, empreintes d'incertitude, il est naturel de croire que l'avenir est ouvert, que nous faisons face à de véritables alternatives et que la décision nous appartient, ultimement. Dans de telles situations, il est naturel de se croire libre. C'est cette capacité qu'auraient les agents rationnels à déterminer le cours de leurs actions parmi différentes alternatives que l'on nomme « libre arbitre » et c'est aux actions produites de cette façon que l'on réserve généralement le qualificatif « libre ».

Les actions libres se présentent ainsi comme un certain type d'actions sur lesquelles nous avons un contrôle particulier. Il reste maintenant à définir plus précisément le type de contrôle dont il est question. Les options ne manquent pas. Il est possible de définir ce contrôle en faisant appel à la notion de désir du second ordre, de conscience, de rationalité, de causalité indéterministe, et j'en passe. Certains diront par exemple qu'une action est libre lorsqu'elle est produite par un mécanisme qui répond adéquatement aux raisons. D'autres diront plutôt qu'une action est libre lorsqu'elle est causée, de manière indéterminée, par des préférences. L'entreprise de définir le contrôle qu'exigent les actions libres n'est pas simple, mais les motivations qui la sous-tendent sont criantes. Pourquoi prenons-nous la peine de délibérer si nos décisions ne sont pas en notre contrôle ? Pourquoi punissons-nous un criminel si celui-ci ne pouvait pas décider autre chose que de commettre le crime qu'il a commis ? Que vaut une promesse tenue sans la possibilité réelle de la violer ? À quoi bon mettre des efforts dans un projet si le succès de celui-ci n'est pas véritablement entre nos mains ?

Comme le dit Susan Pockett (2013), si le libre arbitre n'existait pas, il faudrait l'inventer. Pourtant, l'idée n'est pas aussi absurde qu'elle y paraît. De plus en plus, des philosophes se tournent vers le scepticisme à propos du libre arbitre. Et si le libre arbitre était seulement une nécessité psychologique ou une invention de l'esprit ? Et si le libre arbitre était une illusion ? Si Pockett elle-même ne croit pas que le libre arbitre est une illusion, d'autres le croient. C'est le cas de Saul Smilansky (2000), qui soutient que le libre arbitre est une illusion nécessaire au maintien des sociétés actuelles.

Mais qu'est-ce qui motive certains penseurs à soutenir que le libre arbitre est une illusion ? Certaines menaces envers le libre arbitre nous viennent directement des sciences. Déjà, l'existence du libre arbitre cadre très mal avec le paradigme déterministe des sciences physiques. En effet, si le déterminisme est vrai, il semble que le futur soit fixé d'avance et qu'il n'y ait pas réellement de possibilités alternatives lorsque nous délibérons. Qui plus est, les moments d'indétermination postulés par la physique quantique au niveau subatomique semblent beaucoup trop chaotiques pour être contrôlés par quiconque. De plus, les neurosciences et les sciences cognitives comprennent de mieux en mieux le fonctionnement du cerveau et les facteurs génétiques et environnementaux qui sous-tendent les processus décisionnels et le comportement humain. Ces découvertes laissent peu de place, semble-t-il, pour l'existence d'une capacité de contrôle unique aux agents rationnels.

Tous ne reconnaissent pas les menaces que posent le déterminisme et les découvertes provenant des neurosciences et des sciences cognitives. Certains penseurs soutiennent que le contrôle particulier qu'exigent les actions libres est tout à fait compatible avec le déterminisme et qu'il est exagéré d'accorder autant de poids aux découvertes, importantes, mais fragmentaires, des neurosciences et des sciences cognitives. Ces défenseurs du libre arbitre adoptent ainsi une approche compatibiliste à propos du libre arbitre. D'autres défenseurs du libre arbitre admettent cela n'a pas de sens de parler d'actions libres dans un monde où toutes nos actions sont déterminées par des facteurs génétiques et environnementaux. Ces penseurs affirment du même souffle que le paradigme déterministe des sciences physiques est faux et qu'une meilleure compréhension du cerveau nous prouvera que nos décisions ne sont pas entièrement déterminées. Ces défenseurs du libre arbitre adoptent ainsi une approche libertarienne et développent des modèles indéterministes de libre arbitre.

La littérature philosophique contemporaine est encore aujourd'hui largement dominée par un plaidoyer en faveur du libre arbitre. Les sceptiques à propos du libre arbitre sont de plus en plus nombreux, mais leurs arguments sont quelquefois ridiculisés, et souvent réduits à leur valeur dialectique, permettant uniquement aux non-sceptiques de mieux développer leur propre modèle. Le compatibiliste Daniel Dennett (2013), par exemple, lorsqu'on lui demande ce qu'il pense des positions sceptiques, déclare simplement qu'elles sont toutes sans espoir. Contrairement à Dennett, je ne crois pas que le scepticisme est sans espoir. Bien au contraire, je pense qu'il s'agit d'une avenue prometteuse, convaincante et qui mérite d'être considérée sérieusement en philosophie. Dans ce mémoire, je proposerai ainsi une défense du scepticisme à propos du libre arbitre. Cette défense exigera une évaluation des diverses significations accordées à la notion de libre arbitre et une critique des modèles compatibilistes et libertarianistes de libre arbitre.

Dans le premier chapitre de ce mémoire, je clarifierai les diverses significations accordées à la notion de libre arbitre et je sélectionnerai les modèles non sceptiques les plus prometteurs de la littérature philosophique contemporaine, et ce dans le but d'offrir une critique charitable du libre arbitre. Pour ce faire, je commencerai par proposer une définition générale de ce qu'est le libre arbitre. Nous aurons l'occasion, dans cette section, de distinguer le libre arbitre de la liberté et de la responsabilité morale, d'évaluer deux visions de la causalité et de critiquer deux notions communes de libre arbitre. Par la suite, je présenterai l'approche libertarianiste du libre arbitre. Les modèles libertarianistes sont, je crois, les meilleurs candidats pour une défense du libre arbitre. Il ne me sera pas possible, évidemment, d'examiner tous les modèles libertarianistes de la littérature. En me basant sur les clarifications que j'aurai effectuées plus tôt, je présenterai ce que je crois être l'un des meilleurs modèles libertarianistes, celui de Laura Ekstrom (2000). Finalement, je présenterai l'approche compatibiliste du libre arbitre. Il existe une très grande variété de modèles compatibilistes, et pour rendre compte de cette diversité, j'exposerai deux modèles très respectés mais très différents de compatibilisme, soit celui d'Harry Frankfurt (1969) et celui de John Martin Fischer et Mark Ravizza (1998).

Dans le deuxième chapitre de ce mémoire, j'arguerai qu'il faut rejeter les modèles compatibilistes. Ces derniers, pour permettre l'existence du libre arbitre dans un monde déterministe, diminuent considérablement nos exigences à propos du contrôle particulier qu'exigent les actions libres, en affirmant, par exemple, que ce contrôle n'exige pas

l'existence d'authentiques possibilités alternatives. Je soutiendrai que les compatibilistes, en diminuant leurs exigences, s'éloignent trop de nos intuitions à propos du libre arbitre. Pour ce faire, je ferai appel à l'argument de la conséquence de Peter van Inwagen (1986) et à l'argument de la manipulation Derk Pereboom (2013). À l'aide de l'argument de la conséquence, je tenterai de démontrer que le libre arbitre n'est pas compatible avec le déterminisme et que les compatibilistes, en prétendant le contraire, sont dans l'erreur. À l'aide de l'argument de la manipulation, j'essaierai de montrer que les compatibilistes ne réussissent pas à rendre compte de façon satisfaisante des cas de manipulation et que cela justifie un rejet de leurs modèles. Je ne prétends pas, dans ce chapitre, miner totalement le projet compatibiliste. En me limitant à l'évaluation de seulement deux de leurs modèles, aussi populaires et respectés soient-ils, je mets une grande partie de la littérature de côté. Je suis pleinement conscient de ces limites. Je crois, par contre, que mes arguments sont suffisamment généraux et suffisamment convaincants pour mettre le fardeau de la preuve sur les épaules des compatibilistes. En l'absence de réponses convaincantes à ces deux arguments, je crois qu'il faut préférer l'approche sceptique à l'approche compatibiliste.

Dans le troisième chapitre de ce mémoire, j'arguerai que le libre arbitre libertarien n'existe pas véritablement. Les libertariens exigent la présence d'une indétermination dans l'histoire causale des actions. J'attaquerai cette approche sur deux fronts. Premièrement, j'essaierai de montrer que l'indéterminisme exigé par les libertariens génère de la chance, rendant impossible l'existence d'actions libres. Pour ce faire, je ferai appel à l'argument de la chance de van Inwagen (2000). Deuxièmement, je soutiendrai que la réalisation neuropsychologique des modèles libertariens est impossible en raison de la vérité du déterminisme, ou plus précisément, en raison de la vérité du quasi-déterminisme (Honderich 2002). Encore une fois, je ne prétends pas, en évaluant un seul modèle libertarien, que le libertarisme est éternellement voué à l'échec. Je crois toutefois que les arguments avancés dans ce chapitre diminuent grandement la plausibilité du projet libertarien.

Dans le quatrième chapitre de ce mémoire, je prendrai la défense d'un modèle sceptique de libre arbitre, l'illusionnisme. En effet, à ce point de mon exposé, il apparaîtra que le libre arbitre ne trouve sa place ni dans un monde déterministe ni dans un monde indéterministe. La voie sera enfin ouverte pour une défense du scepticisme. Afin d'exemplifier ce type de modèle, j'exposerai l'illusionnisme de Smilansky (2000). Selon Smilansky, la forme la plus robuste de libre arbitre, à savoir le libre arbitre libertarien, est une illusion normativement

nécessaire, qui permet de combler les lacunes du compatibilisme et du déterminisme dur. J'examinerai ainsi le modèle de Smilansky et soutiendrai qu'il est préférable d'adopter une approche illusionniste à propos du libre arbitre.

# Chapitre 1 : Qu'est-ce que le libre arbitre ?

## 1.0 Introduction

Le problème du libre arbitre est avant tout un problème d'ordre conceptuel. Avant de savoir si le libre arbitre existe, s'il est compatible avec le déterminisme ou s'il est nécessaire à la responsabilité morale, encore faut-il savoir ce que veut dire « être libre ». La tâche n'est pas mince. Cela fait plus de deux mille ans que les philosophes travaillent sur cette question, et les consensus sont rares, voire inexistants. Pendant que les modèles se raffinent, que les notions de libre arbitre se précisent, une thèse gagne en popularité : le scepticisme. Le libre arbitre n'existerait pas. Les scientifiques en sont convaincus, mais les philosophes en doutent. Ces derniers ont raison de douter : on ne se débarrasse pas du libre arbitre aussi facilement. Face aux sceptiques, les défenseurs du libre arbitre n'ont qu'à poser une question : « Qu'est-ce qui n'existe pas, exactement ? » L'embarras ne se fera pas attendre. Personne n'échappe à l'analyse conceptuelle que sous-tend n'importe quelle investigation sérieuse sur le libre arbitre. Qu'entend-on par « libre arbitre » ? Quelles sont les grandes orientations que prennent les défenseurs du libre arbitre ? Comment ces derniers développent-ils leur modèle ? Telles sont les questions auxquelles nous nous intéresserons dans ce chapitre. Afin d'offrir une défense charitable du scepticisme à propos du libre arbitre, il est nécessaire de présenter clairement à quelles thèses nous nous attaquerons. Ainsi, la fonction spécifique de ce chapitre, dans le cadre général de ce mémoire, est double : clarifier les différentes significations accordées à la notion de libre arbitre et sélectionner les modèles les plus prometteurs pour une défense du libre arbitre. Nous serons, par la suite, en mesure de présenter une critique charitable du libre arbitre. Pour remplir les objectifs de ce chapitre, je procéderai en trois temps. Premièrement, j'exposerai une définition générale du libre arbitre. La fonction de cette première section est de mieux circonscrire la notion de libre arbitre et de poser des bases solides aux développements de bons modèles. Deuxièmement, je présenterai ce que je crois être l'avenue la plus prometteuse pour une défense du libre arbitre : le libertarianisme. L'objectif sera de mettre en évidence les atouts du libertarianisme et les défis auxquels il fait face. La plus grande partie de cette section sera réservée à l'examen du modèle de Laura Ekstrom (2000), un modèle exemplaire de libertarianisme. Troisièmement, je présenterai la deuxième grande famille de modèles de libre arbitre, le compatibilisme. Encore une fois, l'objectif sera de mettre en évidence les atouts et les défis de ce type de modèle. Pour exemplifier le com-

patibilisme, j'examinerai les modèles de Harry Frankfurt (1971) et de John Martin Fischer et Mark Ravizza (1998).

## 1.1 Définition générale

La notion de libre arbitre est équivoque. Avant de préciser les différentes significations de la notion de libre arbitre qui seront attaquées dans ce mémoire, il serait souhaitable de se donner une définition générale de ce qu'est le libre arbitre. Nous pourrions dire, pour commencer, que le libre arbitre est une capacité que possèdent les agents rationnels et qui leur permet de choisir un type d'action parmi différentes alternatives (O'Connor 2014). Cette définition, même si elle est tirée de la *Stanford Encyclopedia of Philosophy*, ne va cependant pas de soi. On remarque, déjà, qu'elle n'attribue le libre arbitre qu'aux agents rationnels. Or, certains penseurs vont jusqu'à attribuer un libre arbitre aux invertébrés (Brembs 2011)<sup>1</sup>, et il peut sembler contre-intuitif d'attribuer de la rationalité aux invertébrés. De plus, la définition de Timothy O'Connor lie le libre arbitre à une sélection entre différentes alternatives. Encore une fois, cela ne va pas de soi. Frankfurt (1971), par exemple, soutient qu'un agent agit librement en faisant X seulement s'il désire faire X et s'il désire que son désir soit effectif (s'il a un désir du second ordre). Dans ce cas, le libre arbitre n'est pas, de prime abord, lié à une sélection entre différentes alternatives. Force est de constater que la définition donnée par O'Connor, même si elle se voulait générale, n'est pas assez inclusive pour rendre compte de toutes les notions de libre arbitre de la littérature. Qui plus est, nous verrons que les différentes notions de libre arbitre, communes ou techniques, tendent à s'éloigner de celle-ci. Il est douteux qu'une définition qui ferait l'affaire de tous puisse être formulée. Ceci étant dit, même s'il est difficile de s'entendre sur ce qu'est le libre arbitre, il est déjà plus facile de s'entendre sur ce qu'il n'est pas. C'est ce que je tenterai de faire dans les deux prochaines sections.

### 1.1.1 Distinction entre libre arbitre et liberté

Il est important de distinguer le libre arbitre de la liberté, et il existe différentes façons de le faire. Il est possible, par exemple, de poser une distinction entre la liberté positive et la

---

<sup>1</sup> Björn Brembs défend et utilise une notion biologique de libre arbitre. Il cherche du même coup à s'éloigner des notions métaphysiques. Aussi intéressantes soient-elles, il ne sera pas question de notions biologiques de libre arbitre dans ce mémoire, principalement parce qu'elles ne sont pas attaquées par les sceptiques.

liberté négative (Berlin 1969). Le libre arbitre est associé à la liberté positive, car elle demande la présence de quelque chose (un contrôle, une autodétermination, une autoréalisation, etc.), alors que la liberté négative fait plutôt référence à une absence de contraintes. Il est aussi possible de poser une distinction entre la liberté interne et la liberté externe (Carter 2012). Dans cette optique, le libre arbitre correspond à la liberté interne, car il subit la pression de facteurs internes à l'agent. La liberté externe, quant à elle, subit la pression de facteurs externes à l'agent. Elle semble donc être associée à une liberté d'ordre politique, social ou simplement physique. Finalement, il semble adéquat de distinguer le libre arbitre de la liberté de la façon suivante : « être libre », au sens du libre arbitre, est une propriété qui relève de l'esprit, alors que « être libre », au sens de liberté, est une propriété qui relève du corps. Par exemple, on dira qu'Edmond Dantès, lorsqu'il est emprisonné au Château d'If, n'est pas libre (au sens de liberté), parce que les barreaux de sa cellule l'empêchent de quitter le château. D'un autre côté, on dira que l'abbé Faria, lui aussi emprisonné au Château d'If, est libre (au sens de libre arbitre), parce qu'il est capable de penser et de raisonner clairement. La façon particulière dont on lie le libre arbitre et la liberté n'aura pas beaucoup d'importance. Il est simplement important de comprendre que ces deux notions sont très généralement examinées de manière indépendante dans la littérature philosophique contemporaine. Les possibles liens conceptuels entre ces deux notions peuvent, et seront, mis de côté. Ainsi, il ne sera pas question de liberté, mais uniquement de libre arbitre, dans ce mémoire. Afin d'éviter toute ambiguïté, l'adjectif « libre » sera dorénavant utilisé en ne faisant référence qu'au libre arbitre.

### **1.1.2 Distinction entre libre arbitre et responsabilité morale**

Il est de mise, maintenant, de distinguer le libre arbitre de la responsabilité morale. Il est assez commun de définir la responsabilité morale en termes de blâme et de louange: « to be morally responsible for something, say an action, is to be worthy of a particular kind of reaction—praise, blame, or something akin to these—for having performed it » (Eshleman 2014). Ainsi définie, la responsabilité morale semble, conceptuellement du moins, indépendante de la notion de libre arbitre. Historiquement, par contre, le libre arbitre a souvent été considéré comme une condition nécessaire à la responsabilité morale. Telle était la position défendue par Aristote dans l'*Éthique à Nicomaque*. Bien qu'il s'agisse d'une thèse relativement intuitive, il est tout à fait possible de défendre le contraire. Fischer (1995), par exemple, développe une notion de libre arbitre en termes de contrôle régulateur (*regulative control*) qui

n'est pas nécessaire à la responsabilité morale. Ce type de contrôle, qui requiert des possibilités alternatives ouvertes à l'agent, est attaqué de plein fouet par les fameux contre-exemples de Frankfurt (1969), dans lesquels un agent semble moralement responsable de ses actions, malgré qu'il n'ait pas pu agir autrement. J'aborderai cette conception du libre arbitre plus loin dans ce chapitre. Dans ce mémoire, il ne sera pas question (directement, du moins) de responsabilité morale. Je me permets ainsi de rester agnostique sur la question de savoir si le libre arbitre est bel et bien une condition nécessaire à la responsabilité morale. Toutefois, la majorité des penseurs présentés à travers ce mémoire accepte une telle thèse, alors l'emploi des termes « responsabilité morale » ne pourra être évité indéfiniment. Qui plus est, la notion se révélera particulièrement importante dans le chapitre 4, lors de la présentation du modèle illusionniste de Smilansky (2000).

### 1.1.3 Libre arbitre et causalité

Dans cette section, je distinguerai les théories de la causalité des agents (*agent-causal theories*) des théories de la causalité des événements (*event-causal theories*), et ce afin de mettre de côté les notions de libre arbitre qui supposent l'existence d'une causalité des agents. Commençons par distinguer ces deux types de causalité. Les théories de la causalité des événements défendent l'idée que les causes et les effets sont toujours des événements<sup>2</sup>. Il s'agit de la façon standard de comprendre la causalité. Si je dis, par exemple, qu'un astéroïde a causé la destruction de la station spatiale internationale, il ne faut pas comprendre par là que c'est l'astéroïde lui-même qui en est la cause. On dira plutôt que c'est le mouvement de l'astéroïde (un événement) qui a causé la destruction de celle-ci. La même analyse peut se faire pour les actions libres. Ainsi, un tenant de la théorie de la causalité des événements ne dira pas qu'un agent est lui-même la cause d'une action libre. Il dira plutôt que l'action a été causée par des croyances et des désirs de l'agent, les occurrences de croyance et de désir étant des événements. Les théories de la causalité des agents, de leur côté, défendent l'idée que les actions libres sont causées par un agent. Il faut comprendre par là que la cause d'une action libre se trouve à être une *substance*, soit l'agent lui-même, et non un événement. Aussi faut-il mentionner que les théoriciens de la causalité de l'agent ne

---

<sup>2</sup> Les événements sont des choses qui *arrivent*, contrairement aux objets, qui *subsistent*. De plus, les événements ont nécessairement une temporalité, ils ont (généralement) un début et une fin dans le temps. Pour en savoir plus sur la distinction entre événement et objet, voir (Casati et Varzi 2014).

nient pas la causalité des événements. Ils postulent plutôt deux types de causalité, l'un pour les événements et l'autre pour les agents. L'analyse d'O'Connor (1995) nous permettra d'y voir plus clair. Selon O'Connor, le pouvoir causal d'un objet provient de l'instanciation de certaines propriétés. Par exemple, si un astéroïde peut causer la destruction de la station spatiale internationale, c'est parce qu'il instancie la propriété de se mouvoir à une certaine vitesse. Dans le même ordre d'idée, si un agent peut causer une action libre, c'est parce qu'il instancie certaines propriétés agentielles. O'Connor, par exemple, parle d'une « *volition-enabling property* » (O'Connor 1995, p. 178). L'important, ici, est de comprendre que selon les théoriciens de la causalité de l'agent, le pouvoir causal n'est pas directement lié aux propriétés des événements, il est plutôt lié à certaines propriétés très particulières des agents.

Pour mieux saisir où se situe l'enjeu, mettons-nous à la place d'un théoricien de la causalité de l'agent et acceptons l'idée que le pouvoir causal est lié, d'une façon ou d'une autre, à certaines propriétés particulières de celui-ci. Comment les théoriciens de la causalité de l'agent rendent-ils compte de la causalité des événements ? Selon eux, les événements ont un pouvoir causal du fait qu'un événement est l'instanciation d'une propriété *dynamique* par un objet, c'est-à-dire une propriété qui implique du mouvement. Sur ce point, tout le monde semble être d'accord. En effet, tout le monde accepte qu'un objet *en mouvement* possède un pouvoir causal en vertu du fait qu'il est en mouvement. L'enjeu n'est pas là. La question qu'il faut se poser est plutôt la suivante : est-ce que l'instanciation d'une propriété non dynamique par un objet confère un pouvoir causal à cet objet ? À cette question, les théoriciens de la causalité des événements répondent négativement : le pouvoir causal est lié uniquement aux propriétés dynamiques. C'est ce que veulent dire les théoriciens de la causalité des événements, je crois, lorsqu'ils affirment que les causes et les effets sont toujours des événements. Les théoriciens de la causalité des agents répondent positivement à la question : le pouvoir causal peut être lié à des propriétés non dynamiques. Plus précisément, le pouvoir causal peut être lié à certaines propriétés non dynamiques bien spécifiques, soit aux propriétés *agentielles*, comme la « *volition-enabling property* » d'O'Connor, par exemple. C'est entre autres ce que veulent dire les théoriciens de la causalité de l'agent lorsqu'ils disent que les actions libres sont causées par un agent en tant que substance.

Faisons un pas en arrière. La causalité est généralement considérée comme une relation entre des événements, et non entre des substances non causées et des événements. Ou encore, pour reprendre la distinction utilisée au paragraphe précédent, la causalité est

généralement considérée comme une relation entre des propriétés dynamiques, et non entre des propriétés non dynamiques et des propriétés dynamiques. Bien qu'il existe encore aujourd'hui des défenseurs de la causalité des agents (Chisholm 1976; O'Connor 2005; Rowe 2006; Griffith 2007; Mawson 2011; Steward 2012), la majorité des défenseurs du libre arbitre rejette l'existence de ce type de causalité. La raison peut être exprimée ainsi : la causalité de l'agent semble violer le principe de la fermeture du physique. Ce principe est exprimé par la fameuse maxime « rien ne se perd, rien ne se crée, tout se transforme ». L'idée générale, c'est que l'univers physique contient une quantité fixe d'énergie et qu'il n'est pas possible d'en introduire ou d'en perdre, l'énergie ne fait que se déplacer. Or, si un objet « immobile » (ayant uniquement des propriétés non dynamiques) cause un mouvement, on introduit de l'énergie dans notre système physique. On viole ainsi le principe de la fermeture du physique, un principe qui est au cœur de notre compréhension du monde. Les notions techniques de libre arbitre qui ont été retenues dans ce mémoire ne supposeront pas l'existence d'une causalité des agents. N'ayant pas présenté une analyse systématique de la causalité des agents, je ne prétends pas avoir miné le fondement de ces théories. Par contre, considérant qu'il n'est pas possible d'attaquer toutes les notions de libre arbitre de la littérature à la fois, je me dois de faire une sélection parmi celles qui me semble les plus plausibles, et celles qui supposent l'existence d'une causalité des agents me semble beaucoup plus difficile à défendre<sup>3</sup>.

#### **1.1.4 Les notions du sens commun : présentation et critique**

Avant d'entreprendre l'examen des notions techniques de libre arbitre, il est intéressant de se pencher sur les notions du sens commun. Je fais référence, ici, aux notions qui sont utilisées par la population en général, à ce que l'homme ordinaire *veut dire* lorsqu'il parle de libre arbitre. Évidemment, on peut s'attendre à des notions ambiguës ou partielles. Malgré tout, il est possible de trouver certains éléments qui reviennent plus souvent que d'autres. Dans cette section, je m'intéresserai ainsi à deux notions communes de libre arbitre qui ont été sévèrement critiquées par les sceptiques, tant par des philosophes que par des scientifiques. La première de ces notions est contre-causale et la deuxième fait appel à l'initiation consciente de nos actions. Après avoir examiné chacune de ces notions, je montrerai en quoi

---

<sup>3</sup> Je reviendrai sur la notion de causalité des agents au chapitre 3.

nous sommes justifiés d'être sceptiques à leur sujet. Nous pourrions, par la suite, nous concentrer sur les notions techniques.

L'élément le plus marquant, et sûrement aussi le plus suspect, de la notion commune de libre arbitre est son aspect contre-causal. Le libre arbitre, ainsi conçu, se présente comme une capacité qu'auraient les agents libres à causer certains événements, sans être eux-mêmes causés par quoi que ce soit. En d'autres mots, nous serions tous de petits moteurs non mus (Caruso 2013, p. 5). Aussi étrange que cela puisse paraître, il semble s'agir d'une conception très répandue dans la population générale. Dans un sondage réalisé par Thomas Nadelhoffer et Daniela Goya-Tocchetto, 46% des personnes interrogées sont d'accord avec l'affirmation suivante: « To have free will is to be able to cause things to happen in the world without at the same time being caused to make those things happen. » (Nadelhoffer and Goya-Tocchetto 2013) La grande popularité de cette notion de libre arbitre (en supposant que le sondage de Nadelhoffer et Goya-Tocchetto est fiable) fait en sorte qu'il mérite qu'on s'y attarde un peu. Examinons rapidement quelques raisons qui nous laissent croire que les modèles de libre arbitre fondés sur cette notion ne sont pas valides.

Il apparaît clair que ce type de libre arbitre suppose une causalité des agents. Ce type de causalité, mais cela a déjà été dit plus tôt, viole le principe de la fermeture du physique. Le libre arbitre, ainsi conçu, est donc inconsistant avec les meilleures théories physiques (Pereboom 2013). En effet, il est raisonnable de croire que nos choix sont le produit de certains événements dans notre cerveau. Or, ces événements, si on se fie aux théories physiques, sont gouvernés par des lois naturelles, des lois qui respectent le principe de la fermeture du physique. Ainsi, en postulant que certains agents sont capables de causer certains événements sans être eux-mêmes causés par quoi que ce soit, nous nous retrouvons à dire que certains agents, lorsqu'ils agissent librement, échappent aux lois naturelles, ce qui est évidemment problématique. De plus, le libre arbitre contre-causal soulève des difficultés explicatives concernant le moment d'occurrence (*timing*) des événements causés par un agent (Ekstrom 2000, p. 97). Imaginez qu'un agent prenne une décision au moment  $t_2$ . Pourquoi cette décision arrive-t-elle exactement à  $t_2$ ? Selon les théoriciens de la causalité de l'agent, si l'agent peut causer une décision, c'est parce qu'il instancie une propriété agentielle. Or, l'agent instancie ladite propriété bien avant que la décision survienne. En d'autres mots, l'agent est *capable* de produire des décisions avant qu'une décision se produise effectivement. La capacité de l'agent à produire des décisions ne peut donc pas expliquer pourquoi la

décision s'est produite précisément au moment  $t_2$ . Peut-être la décision s'est-elle produite au moment  $t_2$  parce que l'agent a fait un effort de volonté à  $t_1$ . Mais alors, comment expliquer le moment d'occurrence de l'effort de volonté ? Le même problème se pose à nouveau. Si l'effort de volonté est un événement, nous n'avons plus besoin de faire appel à la propriété agentielle pour expliquer la prise de décision. Nous n'aurons qu'à dire que l'événement « effort de la volonté » a causé l'événement « prise de décision ». Si l'effort de volonté est l'instanciation d'une autre propriété agentielle, nous devons expliquer le moment d'occurrence de l'effort de volonté à  $t_1$ . Les défenseurs du libre arbitre contre-causal semblent ainsi incapables d'expliquer le moment d'occurrence des décisions sans faire appel, ultimement, à autre chose que l'agent lui-même. Pour les deux raisons mentionnées ci-haut, je laisserai de côté la notion contre-causale de libre arbitre, celle-ci faisant face, à mon avis, à de trop grandes difficultés.

Le deuxième aspect de la notion commune de libre arbitre que j'aimerais examiner lie le libre arbitre à l'*initiation consciente* de nos actions. Plus précisément, on dira qu'une action est libre seulement si elle a été initiée de manière consciente par un agent. Cette notion de libre arbitre peut être dégagée des expériences de Benjamin Libet (1999) et de Daniel Wegner (2002). L'invalidité d'une telle notion est moins évidente que celle de libre arbitre contre-causal. Toutefois, si l'on en croit les neurosciences, nous avons de bonnes raisons de douter de l'existence de ce libre arbitre : « Has neuroscience killed free will ? The answer depends entirely on one's definition of free will—and in this case it does not matter whether we are talking about the compatibilist, incompatibilist or libertarian variety. If any of these definition of free will requires the conscious initiation of one action rather than another, then the answer is yes, neuroscience has killed that kind of free will » (Pockett 2013). Afin de comprendre les propos de Pockett, examinons cette fameuse expérience de Libet.

L'expérience de Libet va comme suit : des sujets sont amenés dans une salle où est placée une horloge ne possédant qu'une aiguille. Cette aiguille effectue une rotation complète en quelques secondes. Les sujets, quant à eux, sont branchés à un électroencéphalogramme qui recueille des données sur leur activité cérébrale. On demande alors aux sujets d'accomplir diverses tâches. Dans l'une des expériences, on leur demande de fléchir le doigt lorsque bon leur semble. Dans cette même expérience, les sujets doivent rapporter, à l'aide de la position de l'aiguille sur l'horloge, le moment où ils ont pris conscience de leur intention de fléchir le doigt. L'électroencéphalogramme enregistra une activité cérébrale chez les su-

jets à environ 500 millisecondes *avant* le mouvement de leur doigt. Chez les sujets ayant planifié leurs actes, l'activité cérébrale survient à 1000 millisecondes avant le mouvement de leur doigt. Jusqu'à maintenant, il n'y a rien de bien surprenant. Il est tout à fait normal qu'un signal parte préalablement du cerveau pour commander le mouvement du doigt. Or, n'oublions pas que les sujets avaient aussi comme tâche de rapporter le moment où ils avaient pris conscience de leur intention de fléchir le doigt. À l'aide des données recueillies, on apprend que cette prise de conscience survient généralement à environ 200 millisecondes *avant* le mouvement du doigt, soit environ 300 millisecondes *après* le début de l'activité cérébrale qui initie la flexion. Ainsi, on constate que les activités cérébrales menant à la flexion du doigt commencent avant même que les sujets aient pris conscience de leur intention de fléchir le doigt. Libet conclut, à partir de ces données, que le cerveau initie l'action volontaire de manière inconsciente, puisque des activités cérébrales furent détectées avant que les sujets aient pris conscience de leurs intentions d'agir. Il semble ainsi que le libre arbitre, s'il est défini comme une initiation consciente de nos actions, n'existe pas.

Cette interprétation des résultats de l'expérience de Libet cadre bien dans une perspective épiphénoménaliste à propos de la conscience. L'épiphénoménalisme à propos de la conscience défend l'idée que la conscience ne possède aucun pouvoir causal. Il serait exagéré de soutenir que Libet ait démontré le caractère épiphénoménal de la conscience. Lui-même n'entérine pas cette thèse<sup>4</sup> (Libet 1999). En réalité, l'épiphénoménalisme est, encore aujourd'hui, très controversé (Robinson 2015), et une défense en bonne et due forme de cette thèse nous éloignerait du thème de ce mémoire. Malgré tout, les données recueillies par Libet semblent assez fortes pour nous convaincre qu'il faut abandonner une conception du libre arbitre qui se base sur l'*initiation* consciente de nos actions<sup>5</sup>. En effet, il semble approprié de dire que Libet a minimalement montré que certaines actions que l'on considère comme libres ne sont pas initiées de manière consciente, qu'il ne s'agit là que d'une illusion.

---

<sup>4</sup> Afin de sauver le libre arbitre, Libet postule l'existence d'une capacité de « veto ». Cette capacité, si on en croit Libet, permettrait à un agent de court-circuiter consciemment la chaîne causale menant à l'action. Cette capacité de veto ne semble pas avoir de défenseurs autres que Libet lui-même.

<sup>5</sup> Mise en garde : il n'est pas dit, ici, que la notion de conscience n'est pas liée à celle de libre arbitre, ou que la conscience est un épiphénomène. Par contre, les sceptiques semblent avoir raison de douter que nos actions sont *initiées* consciemment. L'emphase est mise sur le fait que l'*initiation* n'est pas consciente, mais cela ne veut pas dire que la conscience n'a pas un rôle à jouer, ailleurs, durant la chaîne causale. Voir (Mele 2011).

Après tout, si le simple fait de fléchir le doigt n'est pas initié de manière consciente, on peut se demander pourquoi des actions plus complexes le seraient. Même s'il est possible de douter de la validité de l'expérience de Libet (Mele 2007), il faut constater que les notions de libre arbitre faisant appel à la notion de conscience comme initiatrice de l'action ne trouvent aujourd'hui que très peu de sympathisants parmi les philosophes et les scientifiques.

Revenons sur ce qui a été dit jusqu'à maintenant. L'objectif de ce chapitre est de sélectionner les meilleurs modèles que peut offrir la littérature sur le libre arbitre afin d'offrir une réponse sceptique le plus charitable que possible. Avant tout, j'ai commencé par distinguer le libre arbitre de la liberté et de la responsabilité morale, afin de mettre au clair que ce mémoire concerne uniquement le libre arbitre. Par la suite, j'ai défendu l'idée que les notions de libre arbitre qui font appel à la causalité des agents étaient peu plausibles. Ainsi, les modèles qui seront attaqués dans les deux prochains chapitres ne feront pas appel à ce type de causalité. Finalement, j'ai présenté et critiqué deux notions communes de libre arbitre. La notion contre-causale, rejetée par la majorité des philosophes et des scientifiques contemporains (voire la totalité, mais cela demanderait une revue complète de la littérature), est une cible trop facile pour les sceptiques. La notion qui fait appel à l'initiation consciente est aussi à rejeter, car les avancées en neurosciences rendent l'existence d'un tel libre arbitre trop improbable. Ces deux « modèles », jugés non valides, ne seront pas l'objet des arguments sceptiques présentés dans les deux prochains chapitres. À partir de maintenant, je pourrai me concentrer sur des modèles plus techniques de libre arbitre. Ceux-ci sont mieux conçus et seront des « adversaires » de taille pour les sceptiques.

## **1.2 Les notions libertarianistes de libre arbitre**

Ici, je m'intéresserai à une notion de libre arbitre très populaire parmi les philosophes contemporains, soit le libre arbitre libertarianiste. Nullement apparenté à la philosophie politique du même nom, le libertarianisme est la thèse selon laquelle le libre arbitre existe uniquement si le déterminisme est faux. Le libertarianisme est donc une thèse incompatibiliste, c'est-à-dire une thèse selon laquelle l'existence du libre arbitre est incompatible avec la vérité du déterminisme. La contrepartie du libertarianisme est appelée « incompatibilisme dur » (*hard incompatibilism*), la thèse selon laquelle le libre arbitre n'existe pas en raison de la vérité du déterminisme. Le libertarianisme pose ainsi comme condition nécessaire au libre arbitre l'existence d'un moment d'indétermination dans la chaîne causale menant à l'action. Il s'agit,

me direz-vous, d'un pari risqué. Pourquoi aller jusqu'à exiger la fausseté du déterminisme, qui est somme toute une thèse qu'admettent nombre de scientifiques et de philosophes des sciences ? Les libertarianistes partent de l'intuition (et elle est très répandue, si l'on en croit Nadelhoffer et Goya-Tocchetto (2013)) que le libre arbitre ne peut pas exister dans un monde déterministe. Selon les libertarianistes, si l'on veut dire d'un agent qu'il a agi librement, il doit y avoir eu un moment où des alternatives lui étaient ouvertes, une ouverture causale où l'agent a pu déterminer, d'une façon ou d'une autre, quel futur allait être le cas. Dans cette section, je commencerai par présenter ce qu'est le déterminisme, car c'est une thèse qui nous suivra tout au long de ce mémoire. Par la suite, j'examinerai le modèle libertarianiste de Laura Ekstrom (2000). Ce modèle est l'un des modèles les plus respectés et les plus nuancés de libertarianisme. L'examen du modèle d'Ekstrom nous permettra d'exposer clairement les atouts et les défis du libertarianisme. Cet examen se fera de la façon suivante : j'exposerai comment Ekstrom localise le moment d'indétermination et comment, ce faisant, elle évite le problème du contrôle. Finalement, je présenterai les réponses qu'offre Ekstrom à certaines objections que son modèle a reçues.

### 1.2.1 Le déterminisme

Il existe plusieurs types de déterminisme. Dans ce mémoire, je m'intéresserai uniquement au déterminisme nomologique<sup>6</sup>. Le déterminisme nomologique englobe plusieurs types de déterminisme, qu'ils soient physiques, biologiques ou encore génétiques. Tous ces types de déterminisme peuvent constituer une menace à l'existence du libre arbitre (pour les incompatibilistes). Pour l'instant, je m'en tiendrai à une notion générale de déterminisme nomologique. Le déterminisme nomologique peut se définir ainsi : tout événement est *nécessité* par des événements antérieurs en conjonction avec les lois de la nature (Hofer 2010)<sup>7</sup>. Le déterminisme, ainsi défini, est une thèse empirique et contingente à propos des lois de la nature (Vihvelin 2011). Ainsi, le déterminisme sera dit vrai si et seulement si les lois de la nature sont déterministes. Inversement, le déterminisme sera dit faux si et seulement si cer-

---

<sup>6</sup> S'intéresser au déterminisme théologique ou logique nous éloignerait du débat sceptique contemporain, qui s'ancre dans une perspective naturaliste.

<sup>7</sup> Carl Hofer préfère parler de déterminisme causal. Par contre, considérant que sa définition ne contient pas le terme « cause », mais plutôt le terme « lois naturelles », il semble préférable de parler de déterminisme nomologique.

taines lois de la nature sont probabilistes. La loi de la gravitation de Newton, par exemple, peut être dite déterministe<sup>8</sup>, dans la mesure où la gravitation s'applique à tous les corps ayant une masse et qu'il n'est pas possible pour un corps d'avoir une masse sans générer de gravitation. Les lois de Mendel, d'un autre côté, pourraient être dites probabilistes, dans la mesure où l'hérédité génétique possède une composante aléatoire lorsque vient le temps de déterminer quels gènes seront transmis à la descendance. L'entreprise de déterminer si le déterminisme est vrai, ou si le déterminisme s'applique ou non aux comportements humains, sera discutée dans le troisième chapitre.

Dans ce mémoire, les termes « indéterminisme » et « indétermination » seront utilisés pour faire référence à des moments ou à des événements qui ne sont pas nécessaires, en ce sens qu'il *aurait pu en être autrement*. Par exemple, l'effet tunnel (*quantum tunnelling*) est un phénomène indéterministe étudié par la physique quantique qui se produit quand une particule réussit à traverser une barrière et à ressortir de l'autre côté. Admettons que l'effet tunnel puisse s'appliquer à une balle de tennis qui rebondit sur un mur de béton. On dira alors que le rebond de la balle de tennis contre le mur de béton est un événement indéterministe, parce qu'il était possible que la balle, au lieu de rebondir contre le mur, traverse le mur et réapparaisse de l'autre côté. En d'autres mots, il serait possible, dans un monde parfaitement identique au premier, où la même balle de tennis est lancée avec exactement la même force contre un mur exactement identique au premier, que la balle de tennis traverse le mur au lieu de rebondir contre celui-ci. Le rebond de la balle de tennis, n'étant pas un événement nécessaire, sera dit indéterministe.

Il est de mise, maintenant, de distinguer l'indéterminisme ontologique de l'indéterminisme épistémologique. L'effet tunnel était un exemple d'indéterminisme ontologique. Contrairement à l'indéterminisme ontologique, l'indéterminisme épistémologique fait uniquement référence à une incapacité à *prédire* l'arrivée d'un événement. Par exemple, on pourrait soutenir que le moment où, au casino de Charlevoix, la bille noire s'est immobilisée sur la case 24 noire est un moment d'indétermination, mais uniquement au sens épistémologique. En

---

<sup>8</sup> Je ne prétends pas ici soutenir que la loi de la gravitation de Newton est déterministe. Il ne s'agit que d'un exemple hypothétique. Le débat sur le caractère déterministe ou indéterministe de certaines lois particulières m'éloignerait beaucoup trop du propos de ce mémoire. Pour les intéressés, voir (Earman 1986)

effet, il est raisonnable d'affirmer que la force avec laquelle le croupier a lancé la bille noire, en conjonction avec la vitesse de rotation de la roulette, a fait en sorte qu'il était nécessaire que la bille noire s'immobilise sur la case 24 noire. L'événement était donc déterminé au sens ontologique. Par contre, considérant la complexité du calcul pour prédire cet événement, on peut dire qu'il s'agit ici d'un moment d'indétermination au sens épistémologique, parce qu'il n'était pas possible, pour le joueur ou pour le croupier, de prédire que la bille noire s'immobiliserait sur la case 24 noire<sup>9</sup>. Dans le débat sur le libre arbitre, il est généralement question d'indéterminisme ontologique. La raison en est simple. L'indéterminisme épistémologique correspond *grosso modo* à une forme d'ignorance. Or, les libertariens ne cherchent pas à défendre l'idée qu'un individu est libre uniquement s'il ignore ce qu'il va faire. Une telle notion de libre arbitre serait absurde. Ainsi, lorsque j'emploierai les termes « indéterminisme » ou « indétermination », je ferai uniquement référence à l'indéterminisme au sens ontologique.

### 1.2.2 Le libertarisme de Laura Ekstrom

Comment les libertariens conçoivent-ils le libre arbitre et les actions libres ? Nous savons déjà que les libertariens posent comme condition nécessaire la présence d'une indétermination. Ekstrom propose de prendre cette condition comme point de départ: « The construction of an intelligible incompatibilist theory of free will must begin with the libertarian assumption : If we ever act and choose freely, then it is the case that, given the past and the laws of nature, at some moment  $t$  (or at some moments  $t_1, t_2, t_3, \dots$ ), there is more than one physically possible future » (Ekstrom 2000, p. 84). Évidemment, la fausseté du déterminisme ne suffit pas à elle seule à garantir l'existence du libre arbitre. Les libertariens doivent nous expliquer de quelle manière l'indéterminisme rend possible l'existence du libre arbitre. La première étape, semble-t-il, est donc de localiser l'indéterminisme, et c'est ce que tentera de faire Ekstrom. En effet, selon elle, il s'agit là du problème central du libertarisme, de sa difficulté la plus significative. Elle se donne ainsi comme tâche de localiser l'indéterminisme

---

<sup>9</sup> En physique, la notion d'indéterminisme épistémologique est plus forte que cela. Un événement est indéterministe au sens épistémologique lorsqu'il est complètement *impossible* de prédire son arrivée. Dans le cas de la bille noire, une machine très sophistiquée serait capable de prédire l'événement. Le principe d'incertitude d'Heisenberg décrit parfaitement l'indéterminisme épistémologique. Selon ce principe, il est impossible de prédire la position et la vitesse d'un électron à partir d'un certain seuil, parce que la présence d'un instrument de mesure, aussi sophistiqué soit-il, altèrera inévitablement la position ou la vitesse de l'électron.

dans l'historique causal de l'action et d'expliquer pourquoi le lieu de cet indéterminisme est approprié et important. La deuxième étape, selon Ekstrom, est de distinguer les actions libres des simples accidents. En effet, une fois que l'on a localisé correctement le moment adéquat d'indétermination, il faut expliquer comment l'action est suffisamment contrôlée par l'agent pour être dite libre, et non pas accidentelle. Il s'agit là du *problème du contrôle*. Dans cette section, j'examinerai comment Ekstrom prétend, de manière plausible, répondre à ces deux problèmes.

Ekstrom commence par attaquer les thèses selon lesquelles le libre arbitre impliquerait que les actions libres n'ont pas de cause, ou qu'elles sont causées par l'agent en tant que substance. Considérant que ces deux thèses ont déjà été mises de côté précédemment, je ne m'y pencherai pas davantage. Ekstrom développe ainsi une thèse selon laquelle un agent est libre concernant une action particulière uniquement si cette action est causée, de manière indéterminée, par des événements antérieurs dont l'agent est le sujet (Ekstrom 2000, p. 89). Afin de défendre cette thèse, Ekstrom emprunte à G.E.M. Anscombe une distinction entre causalité et nécessité (*necessitation*) (Anscombe 1971). Cette distinction se fonde sur deux considérations. Premièrement, il est possible de reconnaître et de parler de causes sans avoir à reconnaître ou à parler de nécessité. Par exemple, si Dominic a transmis la mononucléose à Jérôme, on dira que le fait que Jérôme ait bu dans la bouteille d'eau de Dominic a causé la transmission de la mononucléose à Jérôme. Par contre, nous n'avons pas besoin d'une loi générale ou de déterminer une nécessité quelconque afin de comprendre qu'il s'agit bel et bien d'une cause. Deuxièmement, Anscombe nous dit qu'une caractéristique commune à toutes les notions de causalité est de dériver un effet d'une cause. Or, les lois de la nature ne nous donnent aucune indication sur cette dérivation. Si un événement A découle d'un événement B, nous ne pouvons pas en conclure que tous les événements semblables à A découleront d'événements semblables à B. Anscombe conclut, avec raison semble-t-il, que la causalité n'est pas identique à la nécessité. Ekstrom entérine cette conclusion. C'est ainsi qu'Ekstrom met de l'avant une distinction entre les causes nécessitantes (ou déterministes) et les causes non nécessitantes (indéterministes). L'idée est simple : une cause nécessitante entraînera toujours un effet qui lui est associé, alors qu'une cause non nécessitante peut échouer à produire son effet. Ekstrom fait ici référence à une notion de causalité probabiliste, notion qui, elle le rappelle, est largement reconnue en philosophie des sciences.

Maintenant que la porte est ouverte pour l'existence d'une action causée de manière indéterminée, Ekstrom entreprend de localiser le moment d'indétermination. Elle considère trois options. Avant de les examiner, mentionnons qu'Ekstrom adhère à la théorie davidsonienne de l'action, selon laquelle les intentions sont des états mentaux distincts des croyances et des désirs. L'historique causal typique d'une action est le suivant : (1) l'agent entre dans un processus de délibération, dans lequel sont pesés certaines de ses croyances et certains ses désirs, jusqu'à la formation d'un jugement ou d'une préférence. (2) L'ensemble complexe de croyances et de désirs cause une intention d'agir. (3) L'intention d'agir cause le comportement associé à l'intention. En gardant cela en tête, trois options s'offrent à Ekstrom. Premièrement, l'indétermination peut se situer lors de la délibération, soit durant (1). Deuxièmement, l'indétermination peut se situer entre le jugement (ou la préférence) et l'intention, soit durant (2). Troisièmement, l'indétermination peut se situer entre l'intention et le comportement, soit durant (3).

Ekstrom ne croit pas qu'il est adéquat que l'indétermination se situe entre l'intention et le comportement. En effet, si tel était le cas, nous manquerions trop de contrôle sur nos actions pour qu'elles soient dites libres. Admettons par exemple que j'ai délibéré pour savoir ce que j'allais manger pour dîner. Après avoir pesé le pour et le contre (de manière déterminée), je forme une préférence pour un sauté au tofu. Cette préférence cause (de manière déterminée) une intention de manger un sauté au tofu. Par contre, considérant que le lien causal entre mon intention et l'action est indéterminé, il y a des chances que mon intention ne cause jamais mon action de manger un sauté au tofu : « you must wait and see whether your act will or will not follow your intention to perform it » (Ekstrom 2000, p. 103). En d'autres mots, placer l'indéterminisme entre l'intention et l'action rendrait la réalisation de l'intention trop aléatoire, hors du contrôle de l'agent. Ekstrom ne croit pas non plus qu'il soit adéquat que l'indéterminisme se situe entre le jugement (ou la préférence) et l'intention. En effet, cette possibilité engendrerait le problème du contrôle, comme dans le cas précédent : « this account has the same problem [...] simply a level up » (Ekstrom 2000 p. 105).

Vous aurez deviné qu'Ekstrom considère donc que l'indéterminisme doit se situer lors de la délibération, c'est-à-dire entre les considérations de l'agent et la formation d'un jugement ou d'une préférence. Avant d'examiner comment Ekstrom évite le problème du contrôle, il est important de comprendre la place centrale qu'occupe la notion de préférence pour l'auteure. Ekstrom définit une préférence comme un certain type de désir soumis à une éva-

luation critique qui prend en considération la conception du « bien » de l'agent (Ekstrom 2000, p. 106). Le contenu de ce désir est qu'un désir du premier ordre, c'est-à-dire un désir portant sur le monde, donne lieu à une action qui le satisfera. Ekstrom fait ici référence à la notion de désir du second ordre d'Harry Frankfurt (1971)<sup>10</sup>. Les actions qui se fondent sur le désir d'un agent ne sont pas nécessairement libres. Ekstrom donne l'exemple des dépendances. Un agent qui consomme de l'héroïne parce qu'il le désire ne sera généralement pas considéré comme libre, à moins qu'il ne possède un désir de désirer de l'héroïne. En plus de permettre d'éviter les contre-exemples de dépendance, les préférences permettent aussi d'éviter les contre-exemples de manipulation. Admettons par exemple qu'Edem, à l'aide de ses superpouvoirs de persuasion, réussit à former dans l'esprit de Sylar le désir de se suicider. Admettons que Sylar, sur la base de son désir, se suicide. On ne pourra pas, selon le modèle d'Ekstrom, dire que Sylar s'est suicidé librement, car son désir n'a pas été soumis à une évaluation critique de sa part, il ne s'agissait pas d'une préférence. Selon Ekstrom, la formation d'une préférence assure qu'il y ait une évaluation critique de la part de l'agent (Frankfurt, quant à lui, n'ira pas jusque là). On parlera d'actions libres, donc, uniquement lorsque celles-ci sont fondées sur un type particulier de désir, soit les préférences.

Évidemment, cela n'est pas suffisant pour un libertarieniste. Comme nous l'avons vu, il faut aussi qu'il y ait un indéterminisme lors de la délibération. Afin de joindre l'idée de préférence à celle d'indéterminisme, Ekstrom développe la notion d'autorisation invaincue (*undefeated authorization*). Cette notion est définie ainsi: « S is authorized in preferring A in a way that is undefeated if and only if S's evaluative faculty was neither coerced nor causally determined by anything to form the preference for A, but rather the preference for A was indeterminately caused by S's considerations » (Ekstrom 2000, p. 108). Par « autorité invaincue », Ekstrom fait référence à la présence d'une causalité non nécessitante entre les considérations d'un agent et la formation de ses préférences. Ainsi, les considérations d'un agent ne déterminent pas complètement la préférence qui sera formée. On dira plutôt que les considérations de l'agent *augmentent la probabilité* qu'une telle préférence soit formée. La causalité non nécessitante, quand elle est localisée lors de la délibération, n'engendre pas, selon Ekstrom, le problème du contrôle. Ekstrom prétend, à l'aide de cette notion, capturer l'essen-

---

<sup>10</sup> Je reviendrai sur les notions de désir du premier et du second ordre à la section sur le compatibilisme de Frankfurt.

tiel de ce qu'est le libre arbitre. Ainsi, le libre arbitre serait une capacité de former des préférences avec une autorisation invaincue. Plus précisément, une action sera dite libre si elle a été causée par une intention qui a elle-même été causée par une préférence avec une autorité invaincue. Un exemple nous aidera à comprendre.

Admettons que Kate, une citoyenne responsable, soit convoquée par le service de sécurité de l'université afin de témoigner à propos d'un événement survenu quelques jours plus tôt. Kate doit ainsi délibérer pour savoir si elle se présentera ou non à la convocation. Plus précisément, sa délibération peut aboutir à trois préférences : la préférence de se présenter et de dire la vérité à l'agent de sécurité, la préférence de se présenter et de mentir à l'agent de sécurité, et la préférence de ne pas se présenter à la convocation. Plusieurs considérations entrent en jeu dans sa délibération. Parmi elles, mentionnons le désir de ne pas se compromettre, le désir de ne pas trahir ses amis, la croyance que cette rencontre est inutile, la pensée que la vérité est préférable au mensonge, la pensée qu'elle a autre chose à faire de sa journée et la croyance que l'agent est sympathique et ne lui veut aucun mal. Chacune de ces considérations dirige rationnellement Kate vers l'une ou l'autre des trois préférences. Admettons maintenant que Kate forme la préférence de se présenter à la convocation et de ne pas dire la vérité, ce qui cause son intention de se présenter à la convocation et de ne pas dire la vérité, ce qui cause l'action de Kate de se présenter effectivement à la convocation et de ne pas dire la vérité. On dira que Kate a agi librement si et seulement si (1) la préférence de Kate n'a pas été forcée (*coerced*) et (2) les considérations en jeu lors de la délibération n'ont pas causé de manière nécessaire sa préférence, c'est-à-dire que les différentes considérations auraient pu aboutir à la formation de n'importe laquelle des trois préférences.

Qu'en est-il du problème du contrôle ? Rappelons-nous qu'Ekstrom rejetait la possibilité que l'indéterminisme se situe entre l'intention et l'action, car cela rendrait l'action accidentelle. Ekstrom rejetait aussi la possibilité que l'indéterminisme se situe entre la préférence et l'intention, car cela aussi rendrait l'intention accidentelle. Dans le même ordre d'idées, il est légitime de défendre l'idée qu'il faut aussi rejeter la possibilité que l'indéterminisme se situe lors de la délibération, car cela rendrait la préférence accidentelle. Afin de répondre à ce problème, Ekstrom développe sa notion d'*agent*. Il est important, en commençant, de préciser qu'Ekstrom défend une notion d'agent qui est ontologiquement et conceptuellement réductible à un ensemble d'événements qu'elle nomme *caractère*. Le caractère d'un agent se veut être la somme de toutes ses préférences et adhésions (*acceptances*). Ainsi, lorsqu'un agent

délibère et forme des préférences, il structure lui-même son caractère. Selon Ekstrom, ce qui est menaçant dans le déterminisme, c'est l'idée que notre caractère est complètement déterminé par les événements passés et les lois naturelles (Ekstrom 2000, p. 115). Or, en localisant un indéterminisme lors de la délibération, Ekstrom met en place l'idée que notre caractère est indéterminé. Par contre, notre caractère, lui, cause de manière déterminée la formation de nos intentions et la réalisation de ces intentions, ce qui nous permet d'éviter le problème du contrôle : « when the appropriately formed preference leads to my act, *what I do is "up to me," since I, quite literally, causally determine it* » (Ekstrom 2000, p. 116). Il semble adéquat, selon moi, de résumer la thèse d'Ekstrom ainsi : nous ne contrôlons pas complètement qui nous sommes, mais nous contrôlons complètement ce que nous faisons.

Examinons comment Ekstrom vante son modèle de libre arbitre. Premièrement, Ekstrom nous dit que son modèle capture nos intuitions sur le libre arbitre, en ce sens qu'il propose une ouverture causale (*causal openness*) dans la détermination du soi (*self*), ce qui fait en sorte que nous ne sommes pas le résultat nécessaire d'événements passés en conjonction avec les lois de la nature. Qui plus est, nos préférences ont l'avantage de pouvoir être expliquées causalement *et* rationnellement. Deuxièmement, Ekstrom nous dit que son modèle est phénoménologiquement exact. En effet, dans le modèle que présente Ekstrom, il existe réellement des possibilités alternatives ouvertes aux agents lorsqu'ils délibèrent, ce qui correspond à ce que nous ressentons lorsque nous délibérons. Troisièmement, le modèle proposé est compatible avec une métaphysique naturaliste : il ne postule pas la présence d'une âme immatérielle, il ne postule ni la causalité de l'agent ni l'absence de causalité et il laisse la porte ouverte à une explication scientifique (Ekstrom 2000, p. 116).

Avant d'examiner les notions compatibilistes de libre arbitre, j'aimerais mentionner certaines objections auxquelles répond Ekstrom. Comment le modèle d'Ekstrom gère-t-il les actions qui ne semblent pas être précédées par une délibération ? Ce matin, par exemple, je suis allé me chercher un café au dépanneur, comme j'ai l'habitude de le faire tous les matins. Or, même s'il semble s'agir d'une action libre, il serait faux de dire que je délibère chaque matin pour savoir si je vais me chercher un café au dépanneur. À cette objection, Ekstrom répond que son modèle ne demande pas qu'une délibération soit présente avant chaque action libre (Ekstrom 2000, p. 117). L'important, dans le modèle d'Ekstrom, c'est que l'action ait été causée par une préférence avec une autorité invaincue. Ainsi, une action peut être dite libre même si la préférence qui lui est associée a été formée deux ans plus tôt. Mon action

d'aller me chercher un café au dépanneur est donc libre, car la préférence d'aller me chercher un café au dépanneur est le résultat d'un processus délibératif (indéterminé) que j'ai eu dans le passé.

Imaginons maintenant qu'un agent délibère et qu'une considération X augmente de 99,9999% la probabilité de causer une préférence Y. Dans cette situation, considérant qu'il y a effectivement un indéterminisme en jeu, l'action qui en résulte sera dite libre par Ekstrom, et ce même si la probabilité de la considération X approche dangereusement 100% (ce qui rendrait l'action non libre). Un tel résultat est-il acceptable ? Selon Ekstrom, il est possible de rendre compte de ce problème en faisant appel à la notion de *degré de liberté* (Ekstrom 2000, p. 125). Par exemple, on dira que les préférences qui ont été causées par des considérations ayant une force causale de plus de 80% engendrent des actions moins libres, car les considérations en jeu avaient peu de chance de faire en sorte qu'il en soit autrement. De la même manière, on dira aussi que les préférences qui ont été causées par des considérations ayant une force causale de moins de 20% engendrent des actions moins libres, car les considérations en jeu avaient trop de chance de faire en sorte qu'il en soit autrement. Admettons, par exemple, que Rosalie délibère pour savoir si elle accepte de suivre une chimiothérapie. Si elle n'accepte pas, ses chances de survie s'en verront grandement diminuées. Rosalie accepte donc de suivre la chimiothérapie. Admettons maintenant que Sophie délibère pour savoir si elle donne 800\$ à une œuvre de charité, et ce en sacrifiant ses vacances d'été au Mexique. Sophie est indécise, mais elle décide finalement de sacrifier ses vacances. Dans le modèle d'Ekstrom, Sophie agit plus librement que Rosalie, car les considérations de survie de Rosalie ont un pouvoir causal beaucoup plus élevé que celles dont tient compte Sophie, sont le pouvoir causal s'approche du 50% (aux fins de l'exemple).

Une dernière objection que nous pouvons soulever contre Ekstrom est que dans son modèle, nous ne contrôlons pas complètement qui nous sommes. Cela pose-t-il problème ? Ekstrom rappelle que nous contrôlons en partie qui nous sommes, car le processus de formation de soi est soumis à notre faculté délibérative ou évaluative (Ekstrom 2000, p. 122). Je défendrai le contraire dans le troisième chapitre de ce mémoire. Ainsi, la présence d'un indéterminisme lors de la délibération ne nous prive pas complètement du contrôle que nous avons sur ce que nous sommes, ou sur ce que nous devenons. Comme il a déjà été dit, les incompatibilistes n'ont pas d'autres choix que de localiser une indétermination quelque part dans la chaîne causale menant à l'action. Or, selon Ekstrom, il s'agit là du meilleur endroit où

localiser celle-ci. Comment l'indétermination d'Ekstrom est-elle réalisée ? En d'autres mots, existe-t-il des preuves d'une telle indétermination lors du processus délibératif ? Pour répondre à cette question, Ekstrom fait référence à des textes de Robert Kane (1996) et de Jesse Hobbs (1991) dans lesquels il est défendu l'idée qu'il existe des systèmes chaotiques dans le cerveau (Ekstrom 2000, p. 124). Ces systèmes chaotiques, justifiés par un paradigme indéterministe en physique quantique, rendraient plausible la présence d'un indéterminisme au niveau cérébral. Je soutiendrai, au chapitre 3, qu'Ekstrom n'échappe pas au problème du contrôle et que la réalisation de son modèle, via les systèmes chaotiques du cerveau, est douteuse. Son modèle, l'un des plus nuancés parmi les libertarianistes, me servira en effet de représentant lorsque viendra le temps d'exposer des arguments sceptiques contre le libertarianisme. Mais avant d'aborder les arguments sceptiques, examinons une seconde façon de défendre le libre arbitre qui est souvent favorisée dans la littérature contemporaine, l'approche compatibiliste.

### **1.3 Les notions compatibilistes de libre arbitre**

Contrairement aux incompatibilistes, les compatibilistes croient que le libre arbitre est compatible avec la vérité du déterminisme. Dans cette optique, les notions compatibilistes de libre arbitre n'exigent pas la fausseté du déterminisme, ce qui peut paraître avantageux dans un monde où le déterminisme fait presque consensus chez les scientifiques. Il est important de mentionner que les compatibilistes ont tendance à lier très étroitement le libre arbitre à la responsabilité morale. Ayant mis de côté précédemment la question de savoir si la responsabilité morale exige le libre arbitre, je tenterai autant que faire se peut de demeurer à l'écart de ce débat. Dans cette section, je commencerai par présenter le compatibilisme d'Harry Frankfurt (1971) selon lequel le libre arbitre est associé à des désirs du second ordre. Par la suite, j'examinerai le compatibilisme de John Martin Fischer et Mark Ravizza (1998), selon lesquels le libre arbitre est plutôt associé à une réponse adéquate face aux raisons. Ces deux modèles compatibilistes de libre arbitre seront les fiers représentants du compatibilisme et feront office de cible aux arguments sceptiques du prochain chapitre.

#### **1.3.1 Le compatibilisme d'Harry Frankfurt**

Frankfurt (1971) développe un modèle compatibiliste qui lie le libre arbitre à la notion de désir du second ordre. Avant d'examiner sa notion de libre arbitre, il est nécessaire d'introduire certaines distinctions. Je commencerai cette section en distinguant les *désirs du*

*premier ordre des désirs du second ordre*. Par la suite, je distinguerai deux types de désirs du premier ordre. L'un d'entre eux, les désirs effectifs du premier ordre, sera associé à la volonté. Ensuite, je distinguerai deux types de désirs du second ordre. L'un d'entre eux, les désirs effectifs du second ordre, sera appelé *volition du second ordre* et sera associé au libre arbitre. Par la suite, je serai en mesure d'exposer une distinction que Frankfurt propose entre la notion de *personne* et celle de « *sanfiche* ». Il sera possible, après cela, d'apprécier la notion de libre arbitre que propose Frankfurt à l'aide de deux exemples, soit celui du drogué consentant et celui du drogué réticent. Nous aurons du même coup l'occasion de constater le caractère compatibiliste du modèle de Frankfurt.

Commençons par distinguer les désirs du premier ordre des désirs du second ordre. Les désirs du premier ordre portent sur le monde<sup>11</sup>. Par exemple « Jérôme désire manger sa quiche » et « Stéphanie désire aller sur la lune » expriment tous les deux des désirs du premier ordre, car ils portent respectivement sur le fait de manger une quiche et sur le fait d'aller sur la lune. Les désirs du second ordre, quant à eux, font référence à des désirs portant sur des désirs. Par exemple, « Serge désire désirer faire du jogging quatre fois par semaine » exprime un désir du second ordre, car il porte sur le désir de faire du jogging quatre fois par semaine. Il est important de comprendre que la présence d'un désir de faire X n'implique pas un désir de désirer faire X, et qu'un désir de désirer faire X n'implique pas non plus un désir de faire X. Par exemple, il semble tout à fait compréhensible de s'imaginer que Serge ne désire pas du tout faire du jogging quatre fois par semaine, mais qu'il désire malgré tout avoir le désir de faire ainsi. De la même manière, il est facile de s'imaginer un individu qui désire fumer une cigarette, mais qui ne désire pas du tout désirer fumer une cigarette. Maintenant que la distinction entre les désirs du premier ordre et les désirs du second ordre est posée, examinons plus particulièrement comment Frankfurt distingue deux types de désirs du premier ordre.

Frankfurt distingue les désirs effectifs du premier ordre des désirs non effectifs du premier ordre. Les désirs effectifs du premier ordre sont des désirs qui motivent un agent à passer à l'action. Les désirs non effectifs du premier ordre sont des désirs qui ne motivent

---

<sup>11</sup> Les désirs portant sur des objets ou des individus n'intéresseront pas Frankfurt. En effet, l'énoncé « Yseult désire Tristan », bien qu'il exprime un désir, n'est pas une attitude propositionnelle.

pas un agent à passer à l'action (ces désirs peuvent avoir une certaine force motivationnelle, mais elle ne sera pas suffisante pour que l'agent passe à l'action). Frankfurt posera une relation d'identité entre les désirs effectifs du premier ordre et la volonté : « the notion of the will, as I am employing it, is not coextensive with the notion of first-order desires. It is not the notion of something that merely inclines an agent in some degree to act in a certain way. Rather, it is the notion of an effective desire – one that moves (or will or would move) a person all the way to action. » (Frankfurt 1971, p. 8). Par exemple, si on reprend les exemples précédents, on pourra dire que « Jérôme désire manger sa quiche » exprime la volonté de Jérôme, en admettant bien sûr que c'est le désir de Jérôme qui a motivé Jérôme à manger sa quiche. Le désir de Jérôme serait donc effectif, car il l'a poussé jusqu'à l'action. D'un autre côté, il est raisonnable de dire que « Stéphanie désire aller sur la lune » n'exprime pas la volonté de Stéphanie, car son désir ne la poussera (probablement) jamais à s'engager dans de longues études en génie aérospatial. Le désir de Stéphanie ne serait donc pas effectif, car il ne l'a pas poussée jusqu'à l'action. La distinction qu'apporte Frankfurt permet ainsi de préciser ce à quoi est associée la volonté, soit les désirs effectifs du premier ordre. Évidemment, c'est ce type de désirs qui sera pertinent pour le modèle de libre arbitre de Frankfurt, car ils sont fortement liés à l'action.

Frankfurt pose ensuite une distinction entre les désirs effectifs du second ordre et les désirs non effectifs du second ordre. Commençons par examiner les désirs non effectifs du second ordre. On dira qu'un agent possède un désir non effectif du second ordre lorsqu'il désire désirer X, mais ne désire pas que son désir de X soit effectif. Frankfurt développe un exemple éclairant à ce sujet. Imaginons qu'un psychiatre ait un client dépendant aux narcotiques. Ce psychiatre croit qu'il sera en meilleure position pour comprendre son client s'il est capable de ressentir l'effet que cela fait que de désirer des narcotiques. Il forme ainsi le désir de désirer des narcotiques. Par contre, Frankfurt soulève avec raison que le psychiatre, malgré son désir de désirer des narcotiques, ne désire probablement pas que son désir soit effectif : « it may be that all he wants is to taste the desire for it » (Frankfurt 1971, p. 9). Ce type de désir du second ordre, parce qu'il a comme objet un désir du premier ordre que l'agent ne désire pas être effectif, ne sera pas pertinent dans le modèle de libre arbitre de Frankfurt, car il n'est pas assez fortement lié à l'action.

Examinons maintenant les désirs effectifs du second ordre. On dira qu'un agent possède un désir effectif du second ordre lorsqu'il désire désirer X et qu'il désire que son désir de

X soit effectif. Par exemple, avec un peu d'introspection, je constate que j'ai le désir de désirer me concentrer sur mon mémoire. Le désir du second ordre a comme objet un désir du premier ordre que je veux être satisfait, il a comme objet une volonté. En d'autres mots, je veux que ce désir me pousse jusqu'à l'action. Contrairement à la situation du psychiatre, où le désir du second ordre n'était pas connecté à un désir du premier ordre que le psychiatre voulait être satisfait, je désire non seulement désirer X, mais je désire que mon désir de X soit effectif, je désire que ce soit ce désir qui me motive. C'est ce type de désirs du second ordre qui intéresse Frankfurt. Il appellera ce type particulier de désirs « volition du second ordre ». Frankfurt considère que les volitions du second ordre sont essentielles au statut de *personne* (Frankfurt 1971, p. 10). Frankfurt appelle les individus dénués de volitions du second ordre des « sanfiches » (*wanton*). Les notions de *personne* et de *sanfiche* ne sont pas primordiales dans le cadre de la discussion qui nous occupe. Par contre, il est important de comprendre que, pour Frankfurt, seules les personnes peuvent avoir un libre arbitre. Ainsi, il est essentiel, pour dire qu'une action est libre, que l'agent qui a produit cette action soit minimalement une personne, c'est-à-dire qu'il soit doté de volitions du second ordre. Un sanfiche, quant à lui, est un individu qui se fiche de sa volonté (Frankfurt 1971, p. 11). Parmi les sanfiches, on retrouve les animaux non humains, les très jeunes enfants et possiblement certains adultes.

Grâce à la notion de volition du second ordre, nous sommes maintenant en mesure de définir plus clairement la notion de libre arbitre de Frankfurt. Selon Frankfurt, une personne agit selon son libre arbitre lorsqu'elle désire ce qu'elle veut désirer, c'est-à-dire lorsque ses volitions du second ordre sont conformes à sa volonté (Frankfurt 1971, p. 15). Pour mieux comprendre l'idée de conformité entre les volitions du second ordre et la volonté, Frankfurt a développé deux exemples, soit celui du drogué réticent (*unwilling addict*) et du drogué consentant (*willing addict*). Imaginez-vous un individu vivant un conflit interne. Il possède à la fois un désir de consommer de la drogue et un désir de ne pas en consommer. Le conflit en est donc un entre des désirs du premier ordre. Cet individu possède aussi une volition du second ordre que le désir de ne pas consommer de la drogue soit effectif. Ainsi, cette personne n'est pas neutre à l'égard du conflit du premier ordre qui se joue en elle. Elle préfère que seul le désir de ne pas consommer de la drogue fasse partie de sa volonté. Malheureusement pour elle, sa dépendance est trop forte et elle se retrouve à consommer de la drogue malgré sa volition du second ordre. Il s'agit du drogué réticent. Le drogué réticent n'agit pas selon son libre arbitre, car sa volonté n'est pas conforme à sa volition du second

ordre. En effet, il possède le désir que le désir de ne pas consommer de la drogue soit effectif, mais c'est le désir de consommer de la drogue qui le pousse à l'action.

Imaginez-vous maintenant un individu vivant exactement le même conflit interne du premier ordre. Par contre, cet individu possède une volition du second ordre que le désir de consommer de la drogue soit effectif. Lui non plus n'est pas neutre à l'égard du conflit qui se joue en lui, mais il préfère que sa volonté soit constituée par le désir de consommer de la drogue. Malheureusement (ou heureusement), sa dépendance est aussi trop forte et il se retrouve à consommer de la drogue, conformément à sa volition du second ordre. Il s'agit du drogué consentant (Frankfurt 1971, p. 19). Le drogué consentant agit selon son libre arbitre lorsqu'il consomme de la drogue, car son action est engendrée par un désir de consommer de la drogue qui est conforme à sa volition du second ordre. Certains trouveront contre-intuitif d'attribuer un libre arbitre au drogué consentant, en ce sens que le drogué consentant était déterminé, par la force de sa dépendance, à consommer de la drogue. Frankfurt rappelle ainsi que cette notion de libre arbitre est neutre par rapport au déterminisme: « It seems conceivable that it should be determined that a person is free to want what he wants to want » (Frankfurt 1971, p. 20). Il ajoute: « it seems conceivable that it should come about by chance that a person is free to have the will he wants » (Frankfurt 1971, p. 20). Nous verrons que le désintérêt de Frankfurt par rapport au déterminisme est généralisé en un désintérêt complet, et problématique, par rapport à l'historique causal menant à la formation des désirs du premier et du second ordre. Cette problématique est réservée pour le prochain chapitre.

La théorie de Frankfurt est considérée comme l'une des formes les plus reconnues de compatibilisme (McKenna et Coates 2009). Il développe un modèle de libre arbitre qui mise essentiellement sur la conformité entre la volonté d'un agent et ses volitions du second ordre. Son modèle est compatible avec la vérité du déterminisme, car il permet d'attribuer un libre arbitre à certaines personnes, et ce peu importe si la chaîne causale menant à leurs actions est déterminée ou indéterminée. Évidemment, la théorie de Frankfurt n'échappe pas aux critiques. Michael McKenna et D. Justin Coates soulèvent, par exemple, que la présence d'un conflit interne du second ordre est problématique dans le modèle de Frankfurt. Que se passerait-il si un agent, en plus d'avoir un conflit à propos de ses désirs, avait un conflit à propos des désirs qu'il veut désirer ? Aussi, il peut sembler problématique que le modèle de Frankfurt mise uniquement sur la conformité entre les volitions du second ordre et la volonté d'un agent, mais ne s'intéresse pas du tout à la formation des volitions du second ordre et des

désirs. En ne s'intéressant pas à la formation des désirs (en d'autres mots, à l'historique causal menant à la formation des désirs), le modèle de Frankfurt s'exposera à l'argument de la manipulation, que j'exposerai également dans le prochain chapitre.

### 1.3.2 Le compatibilisme de John Martin Fischer et Mark Ravizza

John Martin Fischer et Mark Ravizza (1998) développent un modèle compatibiliste de libre arbitre qui mise principalement sur une sensibilité aux raisons. Ainsi, un agent qui ne répond pas adéquatement aux raisons sera dénué de libre arbitre. Afin de comprendre ce modèle, je commencerai par présenter les contre-exemples de Frankfurt au principe des possibilités alternatives, car ceux-ci ont eu une grande influence sur le modèle de Fischer et Ravizza. Le principe des possibilités alternatives nous dit qu'un agent est moralement responsable de ses actions uniquement s'il peut agir autrement. Ainsi, les contre-exemples de Frankfurt mettent de l'avant des situations où un agent semble moralement responsable de ses actions, et ce même s'il ne peut pas agir autrement. Par la suite, je distinguerai le contrôle régulateur du contrôle de guidance et expliquerai le lien que tissent Fischer et Ravizza entre les deux. Le modèle présenté ici se concentre uniquement sur le contrôle de guidance, car c'est seulement ce dernier qui est, selon Fischer et Ravizza, compatible avec le déterminisme. Ce type de contrôle sera ensuite défini en termes de réponses adéquates aux raisons. Après avoir mis tous ces éléments en place, nous serons en mesure d'apprécier la notion compatibiliste de libre arbitre de Fischer et Ravizza.

Les contre-exemples de Frankfurt servent à montrer qu'un individu peut être tenu responsable de ses actions, et ce même s'il ne peut agir autrement (Frankfurt 1969). Ce type de contre-exemple a eu une influence majeure dans le débat sur le libre arbitre, et ce principalement en faveur du compatibilisme. Il a convaincu certains penseurs que le contrôle régulateur, un type de contrôle qui requiert l'existence de possibilités alternatives, n'est pas nécessaire à la responsabilité morale. De ce fait, certains philosophes, comme Fischer et Ravizza, ont concentré leur recherche sur un autre type de contrôle : le contrôle de guidance<sup>12</sup>. Afin

---

<sup>12</sup> Fischer et Ravizza défendent en fait un semi-compatibilisme, la thèse selon laquelle *la responsabilité morale* est compatible avec le déterminisme. En effet, ils considèrent qu'une certaine forme de libre arbitre, le contrôle régulateur, est incompatible avec le déterminisme et qu'une certaine forme de libre arbitre, le contrôle de guidance, est compatible avec le déterminisme. En raison de l'influence des contre-exemples de Frankfurt, Fischer et Ravizza croient que seul le contrôle de guidance est nécessaire à la responsabilité morale.

d'éviter de m'engager dans la question de savoir s'il y a un lien entre le libre arbitre et la responsabilité morale, je considérerai le contrôle de guidance comme une notion de libre arbitre à part entière, malgré le lien étroit qu'il entretient avec la responsabilité morale dans la littérature. Avant d'examiner plus en détail le contrôle de guidance dessiné par Fischer et Ravizza, voici un contre-exemple à la Frankfurt que j'ai conçu. Cela permettra, je l'espère, de comprendre les intuitions qui sont derrière le modèle compatibiliste de Fischer et Ravizza.

Charles a récemment mis en place un plan d'assassinat ayant pour cible Sylvain Bouchard, un animateur controversé de Québec. Charles est dérangé par certains propos sexistes, racistes et homophobes que l'animateur continue d'émettre en ondes chaque semaine. Son plan est le suivant : pendant que Charles divertit la sécurité et ouvre les portes du studio, un complice devra s'introduire dans les locaux où travaille Sylvain Bouchard et loger une balle de revolver dans la tête de celui-ci, ce qui mettra fin à la diffusion des propos insensés de l'animateur. Charles demande donc l'aide de François-Xavier, un ami qui partage les mêmes raisons que Charles de mettre fin à la vie de Sylvain Bouchard. Charles sait que François-Xavier est anxieux. Pour s'assurer que François-Xavier ne change pas d'avis le moment venu, Charles implante, la nuit même, une puce électronique dans le cerveau de François-Xavier (François-Xavier dort d'un sommeil très profond cette nuit-là). Ainsi, si François-Xavier hésite, lorsqu'il fera face à Sylvain Bouchard, la puce enverra une impulsion électrique qui obligera François-Xavier à appuyer sur la détente. Heureusement, François-Xavier ne fait preuve d'aucune hésitation le lendemain, et met fin à la vie de Sylvain Bouchard sans broncher. La puce n'aura servi à rien. Dans cet exemple, il semble très intuitif de rendre François-Xavier responsable de l'assassinat de Sylvain Bouchard, et ce même s'il ne pouvait agir autrement. En effet, l'action de François-Xavier était le résultat d'un processus délibératif normal qui se fondait sur certaines raisons. Ainsi, même si la puce électronique empêchait François-Xavier de ne pas appuyer sur la détente, nous sommes prêts à dire que François-Xavier est responsable du meurtre de Sylvain Bouchard.

Les contre-exemples de Frankfurt, comme celui développé ci-haut, motivent Fischer et Ravizza à mettre de l'avant un type de contrôle qui ne demande pas l'existence de possibilités alternatives (ce pour quoi Ekstrom préfère utiliser l'expression « ouverture causale », comme nous l'avons vu). Ce type de contrôle est appelé « contrôle de guidance ». Avant de développer leur modèle de contrôle de guidance, Fischer et Ravizza clarifient la distinction entre le contrôle régulateur et le contrôle de guidance (Fischer et Ravizza 1998, p. 30). Pour

ce faire, ils font appel à l'exemple d'une automobiliste. Sally, après avoir formé l'intention de tourner à droite, signale son intention à l'aide des clignotants, tourne le volant vers la droite et guide ainsi l'automobile vers la droite. Sally possède un contrôle de guidance sur son automobile, au sens où c'est elle qui a *effectivement* guidé l'automobile vers la droite. Considérant que Sally aurait pu guider l'automobile vers la gauche si elle en avait formé l'intention, on dira qu'elle possède aussi un contrôle régulateur, au sens où Sally avait le pouvoir de guider l'automobile dans n'importe quelle direction. Imaginons maintenant que l'automobile est dotée d'un système de navigation automatique. Ce système de navigation, si Sally avait essayé de tourner à gauche, aurait fait tourner l'automobile vers la droite. Heureusement, Sally avait l'intention de tourner à droite, et c'est ce qu'elle a effectivement fait. Ainsi, dans la situation hypothétique où Sally aurait formé l'intention de tourner à gauche, l'automobile aurait quand même tourné à droite. Dans cette situation, Sally ne possède pas de contrôle régulateur, car elle n'aurait pas pu guider l'automobile dans n'importe quelle direction. Elle possède toutefois un contrôle de guidance, car c'est elle qui a *effectivement* guidé l'automobile vers la droite. Ce que font remarquer Fischer et Ravizza, c'est que le contrôle régulateur, absent dans l'exemple du navigateur automatique, requiert de Sally qu'elle soit capable de guider l'automobile dans la situation actuelle *et* dans une situation alternative si son intention avait été de tourner dans une autre direction. Le contrôle régulateur semble donc n'être, au fond, qu'un double pouvoir d'exercer un contrôle de guidance, c'est-à-dire d'exercer un contrôle de guidance dans la situation actuelle et d'exercer un contrôle de guidance dans une situation alternative (Fischer et Ravizza 1998, p. 31). Il semble adéquat de dire que le contrôle de guidance est une version plus faible de contrôle régulateur, en ce sens où il ne s'applique qu'au monde *actuel*, contrairement au contrôle régulateur qui s'applique aussi à d'autres mondes possibles. Maintenant que la relation entre le contrôle de guidance et le contrôle régulateur est plus claire, il est temps d'examiner plus en détail le contrôle de guidance.

Fischer et Ravizza définissent le contrôle de guidance ainsi : un agent fait preuve de contrôle de guidance sur une action lorsque le mécanisme qui a *effectivement* produit cette action (1) est le sien et (2) répond adéquatement à des raisons (Fischer et Ravizza 1998, p. 39). La première condition sert à discriminer certains mécanismes comme la puce électronique dans l'exemple de l'assassinat de Sylvain Bouchard et le navigateur automatique dans l'exemple de Sally. Dans les cas où ce serait ces mécanismes qui auraient produit l'action, on ne pourrait pas dire qu'il s'agit d'une action libre, car ces mécanismes ne sont pas ceux de l'agent. Pour l'instant, je me concentrerai sur la condition (2) du modèle de Fischer et Raviz-

za, parce qu'il s'agit de l'aspect le plus connu de leur modèle de libre arbitre. J'aurai l'occasion d'examiner plus en détail la condition (1) de leur modèle dans le prochain chapitre, lorsque Fischer et Ravizza seront confrontés à des exemples sophistiqués de manipulation.

Fischer et Ravizza distinguent une forme forte et une forme faible de réactivité aux raisons (*reason-responsiveness*). Les conditions de la réactivité forte aux raisons sont décrites ainsi : si le mécanisme qui produit l'action devait opérer et qu'il existait des raisons suffisantes d'agir autrement, l'agent (a) reconnaîtrait les raisons en question, (b) choisirait d'agir autrement et (c) et agirait autrement (Fischer et Ravizza 1998, p. 41). Selon Fischer et Ravizza, un échec à effectuer (a) est associé aux psychoses délirantes. Un échec à effectuer (b) est associé à des compulsions, à des névroses phobiques ou encore à la faiblesse de la volonté. Un échec à effectuer (c), finalement, peut être associé à des incapacités physiques ou à la faiblesse de la volonté. Si un agent n'échoue pas à effectuer (a), (b) ou (c), on dira qu'il possède une forte réactivité aux raisons.

Fischer et Ravizza ne croient pas que la réactivité forte aux raisons soit nécessaire au contrôle de guidance. La raison en est qu'ils ne semblent pas vouloir exclure le contrôle de guidance dans certains cas de faiblesse de la volonté. En effet, ils développent plusieurs exemples dans lesquels un agent semble agir librement, et ce même s'il est victime de faiblesse de la volonté. Dans l'un de ces exemples, Jennifer décide d'aller à une partie de basketball même si cela va à l'encontre de son meilleur jugement. En effet, Jennifer doit remettre une publication la même soirée, et elle sait qu'elle ne pourra pas y arriver si elle va à la partie de basketball. Même si elle reconnaît qu'elle a de meilleures raisons de rester chez elle et de travailler, « [i]t seems that Jennifer actually goes to the basketball game freely [...] and yet the actual-sequence mechanism which results in her going is not reason-responsive in the strong sense » (Fischer et Ravizza 1998, p.42). Jennifer ne possède pas de réactivité forte aux raisons parce que Jennifer, même si elle reconnaît les raisons qui se présentent à elle *présentement*, choisit de ne pas travailler sur sa publication. Fischer et Ravizza, voulant que Jennifer soit moralement responsable de ses actions, préfèrent ainsi une notion plus faible de réactivité aux raisons. Cette forme de réactivité aux raisons requiert seulement qu'il *existe un monde possible* dans lequel, si un agent était confronté à des raisons suffisantes d'agir autrement, il reconnaîtrait ces raisons, déciderait d'agir autrement et agirait autrement (Fischer et Ravizza 1998, p. 44). Jennifer possède cette forme de réactivité aux raisons. En effet, il existe un monde possible dans lequel Jennifer décide de ne pas aller à la partie de basket-

ball. Admettons, par exemple, que le coût du billet, pour la partie de basket-ball, est de 1000\$. Avec cette raison en tête (cette raison n'était pas présente dans la situation actuelle), Jennifer aurait probablement décidé de ne pas aller à la partie de basket-ball, et ce même si elle est généralement disposée à la faiblesse de la volonté dans de telles situations. Contrairement à la réactivité forte aux raisons, la réactivité faible aux raisons exige seulement qu'il y ait *un monde possible* dans lequel il existe des raisons qui feront changer d'avis Jennifer. Dans ce monde possible, les raisons qui se présentent à Jennifer n'ont pas à être celles qui se présentent actuellement à elle. À l'aide de la notion faible de réactivité aux raisons et de la condition (1) du paragraphe précédent, Fischer et Ravizza considèrent avoir mis en place les conditions nécessaires et suffisantes au contrôle de guidance<sup>13</sup>.

Nous nous retrouvons avec une notion de libre arbitre intéressante. Si l'on en croit Fischer et Ravizza, une action sera dite libre si et seulement si le mécanisme ayant *effectivement* produit cette action (1) est celui de l'agent et (2) possède une réactivité faible aux raisons, c'est-à-dire si, *dans un monde possible* où un agent est confronté à des raisons suffisantes d'agir autrement, l'agent reconnaît ces raisons, décide d'agir autrement et agit autrement. La notion de libre arbitre présentée ici est compatible avec la vérité du déterminisme. Réexaminons le cas de Sally pour s'en convaincre. Admettons cette fois que Sally ait décidé de tourner à droite parce qu'elle souhaite manger une pizza et que le seul restaurant de pizza de la ville se trouve à droite. Encore une fois, le système de navigation automatique fait en sorte que Sally n'aurait pas pu tourner à gauche, et ce même si elle avait eu des raisons de le faire. Sally est-elle, selon le présent modèle, libre de tourner à droite ? Vraisemblablement, le mécanisme délibératif qui a produit sa décision est le sien et elle possède une réactivité faible aux raisons. En effet, la délibération a été effectuée par Sally (il n'y a aucun malin génie, scientifique fou ou autre agent extérieur qui est intervenu), et dans un monde possible où Sally aurait appris, en lisant le journal, qu'un nouveau restaurant de pizza venait d'ouvrir, Sally aurait reconnu cette raison et aurait tourné à gauche pour essayer le nouveau restaurant de pizza. Tu fais erreur, me direz-vous, le système de navigation automatique l'en

---

<sup>13</sup> En réalité, Fischer et Ravizza vont finalement opter pour une version « modérée » de réactivité aux raisons. Il s'agit, comme son nom l'indique, d'une version plus forte que la réactivité faible aux raisons, mais une version plus faible que la réactivité forte aux raisons. Les nuances entre la réactivité faible et la réactivité modérée ne nous intéresseront pas, essentiellement parce que je ne compte pas présenter de critiques de la réactivité faible aux raisons. Pour en savoir davantage, la distinction est posée dans le chapitre 3 de Fischer et Ravizza (1998).

aurait empêchée ! Certes, il existe un monde possible dans lequel le système de navigation aurait empêché Sally de tourner à gauche, mais ce monde possible n'est pas pertinent ici, car, dans le monde possible le plus voisin du nôtre, ce n'est pas le système de navigation qui a effectivement produit l'action, c'est uniquement la délibération de Sally. En d'autres mots, la présence du système de navigation automatique n'est pas requise dans les mondes possibles que l'on doit considérer, car le système de navigation n'a joué aucun rôle dans le processus qui a effectivement produit l'action. Ainsi, il existe bel et bien un monde possible très voisin du nôtre, *dans lequel le système de navigation était absent*, où la délibération de Sally a amené Sally à reconnaître une raison d'agir autrement, à décider d'agir autrement et à agir autrement. Mais Sally ne savait pas qu'il y avait un nouveau restaurant de pizza qui venait d'ouvrir, me direz-vous cette fois ! En effet, Sally n'a pas lu le journal dans la situation *actuelle*, mais dans le monde possible que l'on doit considérer, nous pouvons imaginer que Sally fait face à *d'autres raisons*, et si Sally réagit adéquatement à ces raisons-là, alors on dira que le mécanisme qui a produit l'action y était sensible. Donc, si l'on en croit le modèle de Fischer et Ravizza, Sally était libre de tourner à droite. On voit mieux, maintenant, pourquoi la vérité du déterminisme n'est pas problématique dans le modèle de Fischer et Ravizza. Il en est ainsi essentiellement parce que leur notion de libre arbitre s'intéresse uniquement au mécanisme ayant *effectivement* produit l'action, et ce en dépit des autres facteurs qui empêcheraient l'action de se réaliser, et qu'elle ne se limite pas aux raisons qui se sont *effectivement* présentées à l'agent. Le paragraphe suivant éclaircira davantage ce dont il est question.

On remarque aisément la présence d'un aspect contrefactuel dans le modèle de Fischer et Ravizza: « The view is not merely that an agent would display herself in some counterfactual situations to be responsive to reasons, but rather that her responsiveness to reasons in some counterfactual situations is evidence that her actual conduct *itself*— the causes giving rise to *it* — is also in response to rational considerations » (McKenna et Coates 2009). Certains verront ici une confusion. D'un côté, le compatibilisme de Fischer et Ravizza nous dit qu'il n'est pas nécessaire, pour être libre, d'avoir pu agir autrement. D'un autre côté, cette thèse compatibiliste nous dit qu'il est nécessaire, pour être libre, qu'un agent réagisse autrement si on lui avait présenté des raisons d'agir autrement. La confusion se dissipe rapidement lorsque l'on met au clair les deux interprétations de l'expression « aurait pu agir autrement » (*could have done otherwise*). L'interprétation forte, celle utilisée généralement par les incompatibilistes, nous dit que « aurait pu agir autrement » requiert l'existence de possibilités alternatives ouvertes à l'agent. Ainsi, l'interprétation forte demande que le futur ne soit

pas fixé par les événements du passé et par les lois de la nature (sans quoi il n'existerait pas d'alternatives *ouvertes* à l'agent). L'interprétation faible de « aurait pu agir autrement », celle utilisée généralement par les compatibilistes, ne requiert pas l'existence de possibilités alternatives ouvertes à l'agent, elle est seulement contrefactuelle :

Since determinism is a thesis about what must happen in the future *given the actual past*, determinism is consistent with the future being different *given a different past*. So the classical compatibilists analyzed any assertion that an agent could have done otherwise as a conditional assertion reporting what an agent would have done under certain counterfactual conditions (McKenna et Coates 2009).

Avec cette distinction en tête, on comprend que le contrôle de guidance utilise une interprétation faible de « aurait pu agir autrement », alors que le contrôle régulateur utilise une interprétation forte (Fischer 2006, p. 106).

Le compatibilisme de Fischer et Ravizza fait toutefois face à un problème de taille. Rappelons que la notion de *mécanisme* est centrale dans la définition du libre arbitre que donnent Fischer et Ravizza. Leur théorie n'est pas appelée « mechanism-based reasons-responsive theory » (McKenna et Coates 2009) sans raisons. En effet, leur modèle de libre arbitre s'intéresse uniquement au mécanisme qui a effectivement produit l'action. Par contre, le modèle de Fischer et Ravizza est très flexible concernant les facteurs qui auraient pu empêcher l'action de se réaliser et concernant les raisons que l'agent aurait pu considérer. Dans ces conditions, comment identifier le fameux mécanisme qui donne lieu à l'action ? Quel mécanisme est *le* mécanisme pertinent pour une certaine action ? Malheureusement, Fischer et Ravizza admettent eux-mêmes ne pas avoir de réponse satisfaisante à donner à cette question: « It is simply a presupposition of this theory as presented here that for each act, there is an intuitively natural mechanism that is appropriately selected as the mechanism that issues in action, for the purposes of assessing guidance control and moral responsibility » (Fischer et Ravizza 1998, p. 47). Le problème de l'individuation des mécanismes, même s'il est important, ne pourra être examiné dans ce mémoire parce qu'il est trop spécifique au modèle de Fischer et Ravizza. Je me concentrerai plutôt sur la prétendue compatibilité du libre arbitre et du déterminisme selon le modèle de Fischer et Ravizza, et sur certains problèmes liés aux cas de manipulation.

## 1.4 Conclusion

L'objectif de ce chapitre était de sélectionner et d'exposer les meilleurs modèles de libre arbitre que peut offrir la littérature sur le sujet. Avant de m'attaquer à cette tâche, j'ai indiqué qu'il existe une grande variété de notions de libre arbitre. Pour clarifier le débat, j'ai distingué le libre arbitre de la liberté. Par la suite, j'ai distingué le libre arbitre de la responsabilité morale. Même si le libre arbitre est souvent considéré comme une condition nécessaire à la responsabilité morale, je me suis permis de mettre cette question entre parenthèses. Pour diverses raisons, j'ai aussi mis de côté les notions de libre arbitre qui demandaient la présence d'une causalité des agents. J'ai ensuite examiné et critiqué deux notions communes de libre arbitre. La notion contre-causale de libre arbitre doit être rejetée de par son inconsistance avec les meilleures théories physiques et les difficultés explicatives qu'elle soulève concernant le moment d'occurrence des prises de décision. La notion de libre arbitre qui fait appel à l'initiation consciente de nos actions ne semble guère mieux. En effet, si l'on en croit l'expérience de Libet et les recherches en neurosciences qui s'en sont suivies, il semble improbable que nos actions soient initiées de manière consciente. Ces précisions apportées, nous nous sommes ensuite tournés vers les notions techniques de libre arbitre. J'ai examiné la notion libertarienne d'Ekstrom. Ekstrom développe sa notion de libre arbitre en proposant qu'une action est libre si elle a été causée par une intention qui a elle-même été causée par une préférence avec une autorité invaincue. L'idée générale qui se dégage de son modèle est qu'il existe une certaine part d'indétermination dans la formation de notre caractère. Cette indétermination, essentielle au modèle d'Ekstrom, ne diminuerait pas notre contrôle sur nos actions, car elle est située avant la formation de nos intentions et le passage à l'action. J'ai ensuite examiné la notion compatibiliste de Frankfurt. Selon Frankfurt, une action est libre si elle a été produite par la volonté et si cette volonté est conforme aux volitions du second ordre de l'agent. Cette condition est très populaire dans la littérature. Pour preuve, même Ekstrom, une incompatibiliste convaincue, emprunte à Frankfurt sa notion de désir du second ordre afin de construire sa notion de préférence. J'ai finalement présenté le modèle compatibiliste de Fischer et Ravizza. Ce modèle lie le libre arbitre à une réactivité faible aux raisons. Cette notion contrefactuelle met en évidence qu'une action, pour être libre, doit être une réponse à des considérations rationnelles. J'ose espérer, dans ce chapitre, avoir éclairci les notions de libre arbitre qui seront attaquées dans la suite de ce mémoire.

Dans le prochain chapitre, je soutiendrai que les modèles compatibilistes de libre arbitre sont voués à l'échec. Contrairement aux prétentions des compatibilistes, il est faux qu'un agent, dans un monde déterministe, soit en contrôle de ses actions. Pour le démontrer, je mettrai de l'avant deux arguments suffisamment généraux pour être appliqués à tous les modèles compatibilistes. Le premier de ces arguments, l'argument de la conséquence, met en évidence que le libre arbitre n'est pas compatible avec le déterminisme, parce que notre manque de contrôle sur les événements passés et sur les lois de la nature engendre un manque de contrôle sur les événements présents et futurs. Le deuxième argument, l'argument de la manipulation, met en évidence que tous les modèles compatibilistes sont contraints d'accepter qu'un agent peut être libre même s'il est victime de manipulation. Sur la base de ces deux arguments, j'inviterai mes lecteurs à rejeter en bloc les modèles compatibilistes de libre arbitre, ceux-ci étant fondamentalement insatisfaisants. Je réserve pour le troisième chapitre de ce mémoire la tâche de montrer que les modèles libertarianistes sont tout aussi problématiques.

## Chapitre 2 : Contre le compatibilisme

### 2.0 Introduction

Dans le chapitre précédent, nous avons examiné deux avenues prometteuses pour une défense du libre arbitre : le libertarianisme et le compatibilisme. Les sceptiques à propos du libre arbitre croient qu'aucune de ces deux théories n'est valide. Commençons par nous attaquer au compatibilisme. Même s'il existe une variété impressionnante de modèles compatibilistes, les compatibilistes ont tous une chose en commun : ils croient que le libre arbitre peut exister dans un monde déterministe. Cette idée à elle seule a fait couler beaucoup d'encre. Nos actions peuvent-elles être légitimement qualifiées de « libres » si elles sont le résultat nécessaire des événements du passé en conjonction avec les lois de la nature ? Répondre positivement à cette question devient difficile en considérant que nous n'avons, de manière plausible, aucun contrôle sur les lois de la nature ou sur les événements du passé. Si nos actions sont le résultat nécessaire de facteurs hors de notre contrôle, le libre arbitre semble impossible. Il ne s'agit pas du seul problème auquel fait face le compatibilisme. En mettant de côté nos intuitions sur la compatibilité du libre arbitre et du déterminisme, il est raisonnable d'affirmer qu'un agent qui est victime de manipulation n'est pas libre. Existe-t-il une différence fondamentale entre un agent qui est victime de manipulation et un agent qui vit dans un monde déterministe ? Dans le premier cas, le comportement de l'agent est déterminé par un manipulateur, et dans le deuxième cas, le comportement de l'agent est déterminé par les lois de la nature. Dans les deux cas, par contre, l'agent est déterminé par des facteurs hors de son contrôle. Les compatibilistes doivent ainsi nous expliquer, à partir de leur modèle, comment il distingue les cas de manipulation des cas de déterminisme « ordinaire ». Mon objectif, dans ce chapitre, se veut donc essentiellement critique. Je tenterai de montrer qu'il faut rejeter les modèles compatibilistes de libre arbitre, car ces derniers sont beaucoup trop éloignés de nos intuitions sur le libre arbitre. Pour ce faire, je présenterai l'argument de la conséquence de Peter van Inwagen (1986) et l'argument de la manipulation de Derk Pereboom (2013). Je montrerai, à l'aide de l'argument de la conséquence, que les compatibilistes se fourvoient en prétendant que le libre arbitre est compatible avec le déterminisme. Finalement, je soutiendrai, à l'aide de l'argument de la manipulation, que les compatibilistes ne réussissent pas à rendre compte de façon adéquate des cas de manipulation. Sur la base de ces deux arguments, j'inviterai mes lecteurs à rejeter les modèles compatibilistes.

## 2.1 L'argument de la conséquence

Si le déterminisme est vrai, nos actions sont les conséquences nécessaires des événements du passé en conjonction avec les lois de la nature. Si l'on avait une description complète des événements passés et une connaissance complète des lois de la nature, il nous serait possible de prédire tous les événements de l'univers, de l'explosion d'une étoile lointaine à l'élection du prochain président des États-Unis, en passant par la décision de Susie de manger une crème glacée le 11 juin 2020 à 19h14. Le libre arbitre peut-il exister dans un monde où chacune de nos actions est la conséquence nécessaire des événements du passé en conjonction avec les lois de la nature ? À cette question, van Inwagen répond négativement : cela est impossible. Van Inwagen donne une première formulation de son argument: « If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. » (van Inwagen 1986, p. 56). *Grosso modo*, l'argument de la conséquence nous dit que si nous vivons dans un monde déterminé, alors notre manque de contrôle sur les événements du passé et sur les lois de la nature engendre un manque de contrôle sur les événements du présent et du futur. On remarque facilement en quoi cet argument constitue une menace directe pour les compatibilistes. En effet, ces derniers défendent l'idée que le libre arbitre peut exister dans un monde complètement déterminé, et c'est précisément ce qu'attaque l'argument de la conséquence. L'objectif, ici, sera de défendre le bien-fondé de l'argument de la conséquence afin de mettre en évidence que le compatibilisme est une entreprise vouée à l'échec : le libre arbitre n'est pas compatible avec le déterminisme et un modèle qui prétend le contraire est un mauvais modèle de libre arbitre. Dans cette section, je commencerai par exposer l'argument de la conséquence de van Inwagen. Par la suite, je présenterai une objection de Thomas McKay et David Johnson (1996). Finalement, je répondrai à l'objection de McKay et Johnson et soutiendrai que l'argument de la conséquence de van Inwagen est bien fondé.

Avant de présenter son argument, van Inwagen prend le temps de mettre en place sa définition du déterminisme : les événements du passé en conjonction avec les lois naturelles

impliquent un seul futur possible<sup>14</sup> (van Inwagen 1986, p. 65). Considérant son caractère très intuitif<sup>15</sup> et le fait que remettre en question sa définition du déterminisme nous éloignerait beaucoup de l'argument de la conséquence, je ne remettrai pas en question la définition de van Inwagen.

Van Inwagen présente trois versions de son argument. La première version est formalisée à l'aide de la logique propositionnelle (van Inwagen 1986, p. 70), la deuxième version est formalisée à l'aide du calcul des prédicats (van Inwagen 1986, p. 83) et la troisième version est formalisée à l'aide de la logique modale (van Inwagen 1986, p. 93). Je présenterai uniquement la troisième version de l'argument de van Inwagen, car c'est elle, à mon avis, qui exprime le plus clairement l'argument de la conséquence. Van Inwagen commence par se doter de deux règles. Les voici :

( $\alpha$ )  $\Box p \vdash Np$

( $\beta$ )  $Np \ \& \ N(p \rightarrow q) \vdash Nq$

Commençons par un peu de symbolisme. «  $p$  » et «  $q$  » expriment simplement des propositions quelconques, «  $\rightarrow$  » exprime une implication matérielle et «  $\&$  » exprime une conjonction. «  $\Box$  » est l'opérateur modal de nécessité<sup>16</sup>. «  $Np$  » doit ici être compris comme un opérateur signifiant « personne n'a de contrôle sur la vérité de  $p$ <sup>17</sup> » et le symbole «  $\vdash$  » signifie « est dérivable de ». Ainsi, la règle ( $\alpha$ ) nous dit qu'à partir d'une proposition  $p$  qui est néces-

---

<sup>14</sup> La définition complète va comme suit : pour chaque moment du temps, il existe une proposition qui exprime l'état du monde à ce moment; si  $p$  et  $q$  sont des propositions qui expriment des états du monde à certains moments, alors la conjonction de  $p$  avec les lois naturelles implique  $q$ . Je considère que la version abrégée du déterminisme donnée par van Inwagen est suffisamment claire.

<sup>15</sup> Ekstrom utilise une définition similaire (Ekstrom 2000, p. 84). En fait, il est assez commun de définir le déterminisme en faisant référence à « un seul futur possible ».

<sup>16</sup> Un énoncé est nécessairement vrai si et seulement s'il est vrai dans tous les mondes possibles.

<sup>17</sup> Plus exactement, van Inwagen définit  $Np$  ainsi : « no one has, or ever had, any choice about whether  $p$  » (van Inwagen 1986, p. 93). Plus tôt dans son texte, van Inwagen utilise l'expression « having control over the truth-value of a proposition » (van Inwagen 1986, p. 67). La *Stanford Encyclopedia of Philosophy* préfère utiliser l'expression « no one has power over a fact » (McKenna et Coates 2009). Afin que mes phrases soient compréhensibles, j'utiliserai l'expression « personne n'a de contrôle sur la vérité de  $p$  » ou encore « personne n'a de contrôle sur le fait que  $p$  ». J'aurai l'occasion de préciser la signification d'un tel opérateur un peu plus loin.

sairement vraie, il est possible de dériver que personne n'a de contrôle sur la vérité de cette proposition  $p$ . Par exemple, il est généralement admis que «  $2 + 2 = 4$  » est nécessairement vrai. À partir de là, la règle ( $\alpha$ ) nous permet de dériver que personne n'a de contrôle sur la vérité de «  $2 + 2 = 4$  ». La règle ( $\beta$ ), quant à elle, nous dit qu'à partir d'un manque de contrôle sur la vérité d'une proposition  $p$ , et d'un manque de contrôle sur la vérité d'une proposition de la forme « si  $p$  alors  $q$  », il est possible de dériver un manque de contrôle sur la vérité de la proposition  $q$ . Par exemple, admettons que je n'ai aucun contrôle sur le fait qu'il pleut présentement et que je n'ai aucun contrôle sur le fait qu'il pleut uniquement s'il y a un haut niveau d'humidité dans l'air, alors la règle ( $\beta$ ) permet de dériver que je n'ai pas de contrôle sur le fait qu'il y ait un haut niveau d'humidité dans l'air.

À l'aide des règles ( $\alpha$ ) et ( $\beta$ ), nous sommes en mesure d'examiner l'argument formel de van Inwagen dont la conclusion est que personne n'a de contrôle sur quoi que ce soit (van Inwagen 1986, p. 95).

- (1)  $\Box((H \& L) \rightarrow P)$
- (2)  $\Box(H \rightarrow (L \rightarrow P))$
- (3)  $N(H \rightarrow (L \rightarrow P))$
- (4)  $NH$
- (5)  $N(L \rightarrow P)$
- (6)  $NL$
- (7)  $NP$

Dans cet argument «  $H$  » exprime une proposition vraie à propos de la totalité des événements du passé, «  $P$  » exprime une proposition vraie à propos d'un quelconque événement<sup>18</sup> et «  $L$  » exprime la conjonction de toutes les lois de la nature. Examinons maintenant les prémisses et les inférences. La prémisse (1) est la définition du déterminisme : nécessairement, la totalité des événements du passé jointe à la totalité des lois de la nature sont les conditions suffisantes à l'occurrence de n'importe quel événement (passé, présent ou futur).

---

<sup>18</sup> Van Inwagen utilise «  $P_o$  » et «  $P$  » au lieu de «  $H$  » et «  $P$  ». J'ai préféré utiliser la notation de la *Stanford Encyclopedia of Philosophy* : « Arguments for Incompatibilism » (Vihvelin 2011).

À partir de (1), il est possible de déduire logiquement (2)<sup>19</sup>. En utilisant la règle ( $\alpha$ ) sur (2), il nous est possible de déduire (3). Van Inwagen introduit ensuite la prémisse (4), selon laquelle personne n'a de contrôle sur les événements du passé. En utilisant la règle ( $\beta$ ) sur (3) et (4), nous pouvons déduire (5). Van Inwagen introduit ensuite la prémisse (6), selon laquelle personne n'a de contrôle sur les lois de la nature. En utilisant la règle ( $\beta$ ) sur (5) et (6), van Inwagen peut déduire (7), la conclusion tant attendue selon laquelle personne n'a de contrôle sur quoi que ce soit.

Le présent argument se fonde sur cinq éléments : la vérité du déterminisme, la vérité de la prémisse (4), la vérité de la prémisse (6), la validité de la règle ( $\alpha$ ) et la validité de la règle ( $\beta$ ). Admettons maintenant, comme le prétendent les compatibilistes, que la conclusion est fautive. Van Inwagen nous dit, par exemple, qu'il semble faux que personne n'ait eu de contrôle sur le fait que Richard Nixon ait reçu un pardon pour les fraudes commises au sein de son administration (van Inwagen 1986, p. 95). En partant de l'hypothèse que la conclusion est fautive, nous sommes maintenant dans l'obligation de rejeter au moins un des éléments de l'argument présenté ci-haut. Lequel faut-il rejeter ? Allons-y par élimination. La prémisse (4) ne semble pas pouvoir être rejetée. En effet, que personne n'ait de contrôle sur les événements du passé est hautement intuitif (cette thèse est appelée « fixité du passé »). En attendant l'invention d'une machine à remonter dans le temps, nous semblons contraints d'accepter (4). La prémisse (6) semble elle aussi très plausible. Il est difficile de voir comment nous pourrions avoir un contrôle sur les lois de la nature (cette thèse est appelée « fixité des lois »). Van Inwagen ne voit pas non plus comment il serait possible de rejeter ( $\alpha$ ). Si tel était le cas, cela reviendrait à dire que certains humains ont un contrôle sur ce qui est nécessairement vrai : « Hardly anyone besides Descartes has been willing to concede such a capability even to God » (van Inwagen 1986, p. 96). Nous nous retrouvons maintenant avec seulement deux possibilités : rejeter le déterminisme ou rejeter la règle ( $\beta$ ). Évidemment, van Inwagen veut nous convaincre qu'il faut rejeter le déterminisme si l'on veut permettre au libre arbitre d'exister. Ainsi, l'argument de la conséquence remplira son objectif, c'est-à-dire de défendre l'idée selon laquelle le libre arbitre est incompatible avec la vérité du déterminisme. Van Inwagen doit maintenant démontrer la validité de la règle ( $\beta$ ).

---

<sup>19</sup> Les deux propositions sont logiquement équivalentes, mais si vous voulez vous en convaincre, n'importe quelle méthode de preuve de logique modale fera l'affaire.

Rappelons-nous la règle ( $\beta$ ). Celle-ci nous dit que si personne n'a de contrôle sur la vérité de  $p$ , et si personne n'a de contrôle sur la vérité de  $p \rightarrow q$ , alors personne n'a de contrôle sur la vérité de  $q$ . Van Inwagen considère lui-même que la validité de ( $\beta$ ) est la prémisse la plus difficile à défendre (van Inwagen 1986, p. 96). Afin de montrer la non-validité de ( $\beta$ ), il suffirait de mettre en place une situation où les prémisses sont vraies et la conclusion fausse<sup>20</sup>. Or, van Inwagen ne réussit pas à s'imaginer une instance de ( $\beta$ ) où les prémisses sont vraies et la conclusion fausse. Van Inwagen donne plusieurs exemples dans lesquels la vérité des prémisses semble entraîner la vérité de la conclusion (van Inwagen 1986, p. 98). L'un d'entre eux se fonde sur les deux prémisses suivantes : (a) personne n'a de contrôle sur le fait que le soleil explosera dans 2000 ans et (b) personne n'a de contrôle sur le fait que l'explosion du soleil dans 2000 ans entraînera la fin de la vie sur terre dans 2000 ans. À l'aide de ( $\beta$ ), il est possible de conclure que personne n'a de contrôle sur le fait que la vie sur terre prendra fin dans 2000 ans. Afin de montrer que l'argument est non valide, il faut rendre la conclusion fausse. Pour ce faire, on peut par exemple soutenir qu'il est possible de prévenir l'explosion du soleil, mais ce faisant, la prémisse (a) devient fausse. On peut aussi soutenir que nous pouvons prévenir la fin de la vie sur terre en installant un immense bouclier solaire autour de la terre. Par contre, en faisant ainsi, on rend la prémisse (b) fausse. Il semble ainsi difficile de rendre la conclusion fausse sans nier du même coup l'une ou l'autre des prémisses. La règle ( $\beta$ ) semble donc valide. Évidemment, l'échec à rendre ( $\beta$ ) non valide dans cet exemple particulier ne suffit pas à montrer la validité de ( $\beta$ ). Ne réussissant pas à trouver de contre-exemples, Van Inwagen lance un défi aux lecteurs : « I cannot help feeling that the reader who makes a serious attempt at constructing a counter-example to ( $\beta$ ) will begin to appreciate, even if he does not come to share, the intuition that I have expressed by saying, "when I carefully consider ( $\beta$ ), it seems to be valid"» (van Inwagen 1986, p. 98).

---

<sup>20</sup> De plus, considérant que l'objectif de l'argument de la conséquence est de montrer que le déterminisme est incompatible avec l'existence du libre arbitre, ladite situation ne doit pas rejeter le déterminisme.

### 2.1.1 Le contre-exemple de McKay et Johnson

Prenant le défi au sérieux, McKay et Johnson (1996) réussirent à mettre en évidence que la conjonction de ( $\alpha$ ) et de ( $\beta$ ) implique la validité de la règle d'agglomération<sup>21</sup>. Cette règle nous dit qu'à partir d'un manque de contrôle sur la vérité de  $p$  et d'un manque de contrôle sur la vérité de  $q$ , il est permis de dériver un manque de contrôle sur la vérité de la conjonction de  $p$  et  $q$ . Par exemple, admettons que personne n'a de contrôle sur le fait qu'il pleut présentement et que personne n'a de contrôle sur le fait qu'il y a du brouillard présentement, il est possible de dériver que personne n'a de contrôle sur le fait qu'il pleut et qu'il y a du brouillard présentement. La règle d'agglomération se formalise ainsi :

Agglomération :  $Np \ \& \ Nq \ \vdash \ N(p \ \& \ q)$

Considérant que ( $\alpha$ ) et ( $\beta$ ) impliquent la règle d'agglomération, il ne reste qu'à trouver un contre-exemple à la règle d'agglomération, et la règle ( $\beta$ ) sera invalidée du même coup. Le contre-exemple de McKay et Johnson va comme suit : admettons qu'un individu ait en sa possession une pièce de monnaie avec laquelle il *n'a pas* joué à pile ou face hier. Supposons maintenant que  $p$  exprime « la pièce de monnaie n'est pas tombée sur pile hier » et que  $q$  exprime « la pièce de monnaie n'est pas tombée sur face hier ». À partir de là, il est facile de soutenir que personne n'a eu de contrôle sur la vérité de l'énoncé « la pièce de monnaie n'est pas tombée sur pile hier » et que personne n'a eu de contrôle sur la vérité de l'énoncé « la pièce de monnaie n'est pas tombée sur face hier ». En effet, même si l'individu avait décidé de jouer à pile ou face, il n'avait, de manière plausible, aucun contrôle sur le fait que la pièce tombe sur pile ou sur le fait que la pièce tombe sur face. Les prémisses semblent donc vraies. Grâce à la règle d'agglomération, il est permis de dériver  $N(p \ \& \ q)$ , c'est-à-dire que personne n'a eu de contrôle sur la vérité de « la pièce de monnaie n'est tombée ni sur pile ni sur face hier ». Or, cette conclusion est fausse. Clairement, la personne aurait pu décider de jouer à pile ou face hier et ce faisant, il aurait été faux que la pièce de monnaie n'est tombée ni sur pile ni sur face hier. Convaincu par la démonstration, van Inwagen s'avoue vaincu : « The case imagined is [...] undeniably a counterexample to Agglomeration. Agglomeration is

---

<sup>21</sup> La démonstration est très simple. En partant de l'hypothèse de  $Np$ , de l'hypothèse de  $Nq$  et de la vérité logique «  $\Box (p \rightarrow (q \rightarrow (p \ \& \ q)))$  », il est possible de déduire  $N(p \ \& \ q)$  en seulement quelques lignes en utilisant les règles ( $\alpha$ ) et ( $\beta$ ). Voir (McKay and Johnson 1996).

therefore invalid, and the invalidity of  $(\beta)$  follows from the invalidity of Agglomeration » (van Inwagen 2000).

### 2.1.2 Sauver la règle $(\beta)$ en précisant la signification de l'opérateur $Np$

Face à ce contre-exemple, van Inwagen croit pouvoir rendre la règle d'agglomération valide en modifiant légèrement la signification de l'opérateur  $Np$ . Pour ce faire, il fait appel à la notion d'*accès à des régions de l'espace logique* et distingue deux interprétations de l'opérateur  $Np$  (van Inwagen 2000, p. 8). Selon l'interprétation faible, l'opérateur  $Np$  signifie «  $p$  et toutes les régions de l'espace logique auxquelles quelqu'un a ou aura accès chevauchent  $p$  ». Selon l'interprétation forte,  $Np$  signifie «  $p$  et toutes les régions de l'espace logique auxquelles quelqu'un a ou aura un accès exact est une sous-région de  $p$  ». La solution exposée par van Inwagen, on le constate aisément, est extrêmement complexe. Ainsi, j'exposerai plutôt une simplification de sa solution.

Que veut-on dire par « personne n'a de contrôle sur la vérité de  $p$  » ? Il est possible de dégager (au moins) deux interprétations de cet énoncé. Selon l'interprétation faible, personne n'a de contrôle sur la vérité de  $p$  si et seulement si personne ne peut agir de telle sorte que  $p$  sera *assurément* faux. C'est en ce sens que la règle de l'agglomération n'est pas valide. En effet, même si personne ne peut agir de telle sorte que la pièce de monnaie tombera *assurément sur pile* ou tombera *assurément sur face*, il est faux que personne ne peut agir de telle sorte que la pièce tombera *assurément sur pile ou face*. En effet, le simple fait de lancer la pièce de monnaie dans les airs nous assure (de manière plausible) qu'elle tombera sur l'une de ses deux faces. Selon l'interprétation forte, personne n'a de contrôle sur la vérité de  $p$  si et seulement si personne ne peut agir de telle sorte que  $p$  puisse être faux. En ce sens, la règle d'agglomération est valide. En effet, en ce sens, les prémisses de l'argument sont fausses : il est faux que personne ne peut agir de telle sorte que la pièce de monnaie puisse tomber sur pile et il est faux que personne ne peut agir de telle sorte que la pièce de monnaie puisse tomber sur face. Il est raisonnable d'affirmer qu'un individu, en lançant la pièce de monnaie, agisse de telle sorte que la pièce puisse tomber sur pile et puisse tomber sur face. Ainsi, en utilisant une interprétation forte de  $Np$ , la règle d'agglomération est valide, et la validité de la règle  $(\beta)$  est conservée.

Avant de déclarer victoire, il est nécessaire de revisiter l'argument de la conséquence à la lumière de l'interprétation forte de  $Np$ . La modification de l'opérateur  $Np$  ne semble pas

avoir d'impact sur la vérité de la prémisse (4) ou sur la vérité de la prémisse (6). En effet, il est raisonnable d'affirmer que personne ne peut agir de telle sorte que les événements du passé puissent ne plus être le cas. Il est aussi raisonnable d'affirmer que personne ne peut agir de telle sorte que les lois de la nature puissent ne plus être le cas. La validité de la règle ( $\alpha$ ), elle aussi, reste intacte. En effet, il est plus qu'intuitif d'affirmer que personne ne peut agir de telle sorte qu'un énoncé nécessairement vrai puisse être faux. Finalement, la règle ( $\beta$ ) nous dit maintenant que si personne ne peut agir de telle sorte que  $p$  puisse être faux, et si personne ne peut agir de telle sorte que  $p \rightarrow q$  puisse être faux, alors personne ne peut agir de telle sorte que  $q$  puisse être faux. Sous cette interprétation forte de l'opérateur  $Np$ , le contre-exemple de McKay et Johnson ne tient plus et ne peut plus servir à attaquer la validité de la règle ( $\beta$ ). En attendant un nouveau contre-exemple, la validité de la règle ( $\beta$ ) demeure très intuitive. Ainsi, il semble qu'en utilisant une interprétation forte de l'opérateur  $Np$ , il est possible de conserver la validité de la règle ( $\beta$ ) sans altérer le fondement de l'argument de la conséquence.

### 2.1.3 Un conflit d'intuitions

Certains compatibilistes, confrontés à l'argument de la conséquence, soutiennent que ( $\beta$ ) n'est pas valide parce que la notion de contrôle qu'ils ont développée est compatible avec la vérité du déterminisme (Foley 1979; Slote 1982). En effet, on pourrait croire que Frankfurt, avec sa notion de libre arbitre basée sur la conformité entre la volonté d'un agent et ses volitions du second ordre, n'est pas menacé par l'argument de la conséquence. De la même façon, Fischer et Ravizza, avec leur notion de contrôle de guidance, pourraient se dire à l'abri de l'argument de la conséquence parce que leur notion de contrôle est compatible avec le déterminisme<sup>22</sup>. Van Inwagen croit ici être en présence d'un « genuine [...] conflict of rock-bottom intuition » entre la plausibilité du compatibilisme et la plausibilité de la règle ( $\beta$ ). Pour régler ce dilemme, van Inwagen fait appel à l'intuition de ses lecteurs : « Rule ( $\beta$ ) seems obviously right and compatibilism does not seem obviously right. If two principles are in conflict and one of them seems obviously right and the other does not seem obviously right, then, if one must choose, one should accept the one that seems obviously right. » (van Inwagen

---

<sup>22</sup> En réalité, il semble que n'importe quel compatibiliste ayant développé un tant soit peu sa notion de contrôle peut formuler ce genre d'objections.

1986, p. 103). Considérant que la validité de ( $\beta$ ) est plus plausible<sup>23</sup> que les notions compatibilistes de libre arbitre, van Inwagen nous invite donc à rejeter les notions compatibilistes de libre arbitre.

L'argument de la conséquence est considéré, encore aujourd'hui, comme le meilleur argument démontrant l'incompatibilité du libre arbitre et du déterminisme (Vihvelin 2011). À l'aide de cet argument, j'ai cherché à montrer que les modèles compatibilistes de libre arbitre sont dans l'erreur en prétendant pouvoir réconcilier le libre arbitre avec le déterminisme. Dans un monde déterminé, nous n'avons aucun contrôle sur quoi que ce soit, et un modèle de libre arbitre qui prétend le contraire est un mauvais modèle de libre arbitre. L'argument de la conséquence tire toute sa force du fait qu'il s'attaque à une idée défendue par tous les compatibilistes, c'est-à-dire l'idée que le déterminisme ne menace pas l'existence du libre arbitre. Grâce à l'argument de la conséquence, je crois avoir montré que l'entreprise compatibiliste est difficilement défendable pour ne pas dire vouée à l'échec. Dans la prochaine section, j'examinerai un deuxième argument contre le compatibilisme, soit l'argument de la manipulation. Je crois que cet argument saura convaincre ceux que l'argument de la conséquence aura épargnés parce que, d'une part, il mise sur une propriété que l'on oppose fortement au libre arbitre, soit la propriété d'être manipulé et que, d'autre part, il est construit de telle sorte qu'il peut cibler des modèles compatibilistes particuliers. En me confrontant à des modèles compatibilistes particuliers, je pourrai exemplifier ma thèse selon laquelle aucun modèle compatibiliste ne réussira jamais à fournir un modèle satisfaisant de libre arbitre.

## 2.2 L'argument de la manipulation

La manipulation, selon le *Larousse*, est l'action d'orienter la conduite de quelqu'un sans qu'il ne s'en rende compte. La manipulation se distingue de la contrainte, une façon beaucoup moins subtile d'orienter la conduite de quelqu'un. Par exemple, si un bandit entre dans une banque, pointe un revolver au visage du caissier et lui ordonne de lui donner le contenu de sa caisse, on dira que le caissier n'a pas agi librement en remettant le contenu de sa caisse au bandit, il a été contraint (ou forcé) d'agir. L'absence de libre arbitre, dans les cas

---

<sup>23</sup> Contrairement à van Inwagen, je ne crois pas que la règle ( $\beta$ ) est *d'évidence* valide. Elle me semble, par contre, beaucoup plus intuitive que les notions compatibilistes. Évidemment, il ne s'agit ici que de *mes* intuitions.

de contrainte, relève presque de l'évidence. Les modèles de libre arbitre doivent, et réussissent assez facilement, à rendre compte des cas de contrainte. Frankfurt, par exemple, peut discriminer les cas de contrainte de la façon suivante : le caissier, lorsqu'il remet l'argent au bandit, n'agit clairement pas en fonction d'un désir de remettre l'argent au bandit, il n'est donc pas libre lorsqu'il remet l'argent au bandit. Les cas de contrainte ne nous intéresseront pas davantage. Les exemples de manipulation sont beaucoup plus intéressants pour tester la force des modèles de libre arbitre. Au lieu d'orienter directement l'action, le manipulateur oriente les états mentaux de l'agent, ce qui va, ultimement, orienter ses actions. Pour être satisfaisant, je crois qu'un modèle de libre arbitre doit aussi pouvoir discriminer les cas de manipulation. Je soutiendrai que les modèles compatibilistes échouent à cette tâche. L'argument de la manipulation peut être formulé de différentes manières. Les formulations les plus populaires sont celles de Robert Kane (1996), de Richard Taylor (1991) et de Derk Pereboom (2001). Pour exposer l'argument de la manipulation, j'utiliserai la formulation de Pereboom (2013), parce qu'elle est la plus populaire et la plus récente d'entre toutes, et parce qu'elle s'en prend explicitement au modèle de Frankfurt et à celui de Fischer et Ravizza exposés dans le premier chapitre de ce mémoire. Dans cette section, je chercherai à montrer que les modèles compatibilistes de libre arbitre sont insatisfaisants, parce qu'ils ne réussissent pas à rendre compte des cas de manipulation de manière adéquate. Pour ce faire, je commencerai par présenter l'argument de la manipulation de Pereboom. Par la suite, je distinguerai deux types de modèles de libre arbitre : les modèles historiques et les modèles non historiques. Les modèles non historiques, comme celui de Frankfurt, sont frappés de plein fouet par l'argument de la manipulation et seront rapidement jugés insatisfaisants. Les modèles historiques, comme celui de Fischer et Ravizza, nécessiteront une analyse plus approfondie. Après avoir développé davantage le modèle historique de Fischer et Ravizza (1998), je serai en mesure d'offrir une critique de ce modèle. Nous constaterons finalement que les modèles compatibilistes de libre arbitre, même lorsqu'ils sont historiques, échouent à répondre de manière convaincante à l'argument de la manipulation. Les modèles compatibilistes de libre arbitre seront donc rejetés en bloc.

L'exemple initial de Pereboom va comme suit : Le Professeur Plum prend la décision d'assassiner Madame White pour ses bénéfices personnels et assassine effectivement White suite à sa décision. Or, une équipe de neuroscientifiques a été capable de manipuler les états mentaux de Plum à l'aide de technologies radio. Avant que Plum ne commence à délibérer à propos de l'assassinat de White, les neuroscientifiques ont appuyé sur un bouton, mettant

Plum dans un état neuronal faisant en sorte que son prochain raisonnement serait fortement égoïste. Si les neuroscientifiques n'étaient pas intervenus, Plum n'aurait pas décidé d'assassiner White, car son état n'était pas assez égoïste sans l'intervention des neuroscientifiques. Ajoutons que le désir effectif du premier ordre d'assassiner White était conforme au désir du second ordre que le désir d'assassiner White soit effectif et que Plum a agi en fonction de ces désirs. Qu'est-ce que cela veut dire ? On se rappellera que selon Frankfurt, une personne agit selon son libre arbitre lorsqu'elle désire ce qu'elle veut désirer. Plus précisément, Frankfurt associe la volonté à des désirs effectifs du premier ordre, c'est-à-dire à des désirs qui motivent à passer à l'action. Dans le cas présent, la volonté de Plum est d'assassiner White. Pour que cette volonté soit libre, la théorie de Frankfurt stipule que l'agent doit avoir un désir du second ordre dont le contenu est que ce désir le pousse à l'action. Selon Frankfurt, la présence d'un désir effectif du second ordre est essentielle pour dire que Plum a été libre. Dans cet exemple, on dira non seulement que Plum avait le désir d'assassiner White, mais on dira aussi qu'il avait le désir que son désir d'assassiner White le pousse à l'action. Les conditions posées par Frankfurt sont donc remplies et Frankfurt est obligé de dire que Plum était libre d'assassiner White.

Fischer et Ravizza semblent aussi obligés d'en arriver à la même conclusion. En effet, le mécanisme qui a produit la décision de Plum, une délibération dans ce cas-ci, semble (1) être celui de Plum et (2) être sensible aux raisons. J'examinerai sous peu comment Fischer et Ravizza développent la première condition et comment, à l'aide de celle-ci, ils prétendent rendre compte des cas de manipulation. Nous pouvons remarquer que Pereboom ne mentionne pas cette première condition dans son exemple. La discussion m'amènera cependant à montrer que la condition (1) est bel et bien respectée dans l'exemple de Pereboom.

Pour l'instant, concentrons-nous sur la deuxième condition du modèle de Fischer et Ravizza. Selon Fischer et Ravizza, un mécanisme est sensible aux raisons si, dans un monde possible où un agent est confronté à des raisons suffisantes d'agir autrement, l'agent reconnaît ces raisons d'agir autrement, décide d'agir autrement et agit autrement. Lorsqu'elle est présente, cette sensibilité aux raisons doit nous amener à conclure que le mécanisme qui a causé l'action est une réponse à des considérations rationnelles (McKenna et Coates 2009). Dans le cas présent, il semble bien que la délibération de Plum était sensible aux raisons. En effet, il est facile d'affirmer que si, contrairement à ce qui s'est effectivement passé, Plum n'avait eu aucun bénéfice à assassiner White et qu'il avait reconnu qu'il n'avait rien à

gagner à assassiner White, alors Plum n'aurait pas décidé d'assassiner White et ne l'aurait pas tuée. Les neuroscientifiques ont rendu Plum très égoïste, certes, mais l'état d'égoïsme de Plum n'aurait pas suffi à motiver l'assassinat de White sans les possibles bénéfiques personnels que Plum convoitait. Les conditions posées par Fischer et Ravizza semblent donc remplies. En somme, Pereboom montre que même si toutes les conditions compatibilistes sont remplies<sup>24</sup>, il ne semble pas que Plum ait été libre d'assassiner White dans ce premier exemple car il ne l'aurait pas fait sans la manipulation des neuroscientifiques.

Le deuxième exemple de Pereboom est très similaire au premier. Encore une fois, Plum décide d'assassiner White et assassine effectivement White suite à sa décision. Cette fois, par contre, les neuroscientifiques ne sont pas intervenus immédiatement avant la délibération de Plum. En fait, Plum a été programmé lorsqu'il était enfant par la même équipe de neuroscientifiques, de telle sorte que les raisonnements de Plum seraient, parfois, fortement égoïstes. Encore une fois, si les neuroscientifiques n'avaient pas programmé Plum lorsqu'il était enfant, Plum n'aurait pas décidé d'assassiner White. Encore ici, le désir effectif du premier ordre d'assassiner White était conforme au désir du second ordre que le désir d'assassiner White soit effectif et Plum a agi en fonction de ces désirs. De plus, le mécanisme qui a produit la décision de Plum (sa délibération) était celui de Plum et était sensible aux raisons. Les compatibilistes peuvent-ils, contrairement à la première situation, soutenir que Plum était libre d'assassiner White ? Selon Pereboom, il n'existe aucun principe pouvant justifier une telle affirmation (Pereboom 2013, p. 23). En effet, toutes les conditions compatibilistes sont remplies et la seule différence avec la première situation est la présence d'un laps de temps plus long entre l'intervention des neuroscientifiques et la prise de décision de Plum. Or, le fait qu'il se soit écoulé deux secondes ou 40 ans entre l'intervention des neuroscientifiques et la prise de décision ne semble rien changer au fait que Plum a été manipulé par l'équipe de neuroscientifiques.

Le troisième exemple de Pereboom conserve les mêmes paramètres que le deuxième. Cette fois, par contre, ce n'est pas une équipe de neuroscientifiques qui a programmé

---

<sup>24</sup> Plum, dans l'exemple de Pereboom, remplit aussi toutes les conditions compatibilistes posées par David Hume (1738), par R. Jay Wallace (1994) et par Alfred Mele (1995). Considérant que je n'ai pas exposé ces modèles compatibilistes, je me suis permis de les exclure de l'exemple. Pereboom prétend que Plum, dans son exemple, remplit les conditions de *tous* les modèles compatibilistes.

Plum à s'engager, parfois, dans des raisonnements fortement égoïstes. C'est plutôt la famille et la communauté de Plum qui a formé Plum à s'engager dans de tels raisonnements. Au moment de sa formation, Plum était trop jeune pour empêcher sa famille et sa communauté de le former ainsi. De plus, si la famille et la communauté de Plum n'avaient pas formé Plum de la sorte lorsqu'il était enfant, Plum n'aurait pas décidé d'assassiner White. Ici aussi, toutes les conditions compatibilistes sont remplies. En effet, le désir effectif du premier ordre d'assassiner White était conforme au désir du second ordre que le désir d'assassiner White soit effectif et le mécanisme qui a produit la décision de Plum était celui de Plum et était sensible aux raisons. Les compatibilistes, cette fois, peuvent-ils soutenir que Plum était libre d'assassiner White ? Encore une fois, Pereboom soutient qu'il n'existe aucun principe pouvant justifier cela. La différence entre le deuxième et le troisième exemple se joue essentiellement au niveau des intervenants. Dans le deuxième exemple, il s'agissait de neuroscientifiques, et dans ce troisième exemple, il s'agit de la famille et de la communauté de Plum. Or, cette différence ne semble pas pouvoir justifier que Plum ait été libre, dans ce cas-ci, d'assassiner White.

Dans l'exemple final de Pereboom, la seule chose que nous devons assumer est la vérité du déterminisme. Ainsi, tous les événements sont le produit nécessaire des événements du passé en conjonction avec les lois de la nature. Plum est un être humain ordinaire, élevé dans des circonstances ordinaires, et ses raisonnements sont parfois fortement égoïstes. Encore une fois, la décision de Plum d'assassiner White remplit toutes les conditions compatibilistes. Selon quel principe les compatibilistes peuvent-ils justifier que Plum, contrairement au troisième exemple, ait été libre d'assassiner White ? Pereboom considère, encore une fois, qu'un tel principe n'existe pas. Le premier réflexe des compatibilistes sera, semble-t-il, de défendre l'idée que la décision de Plum, dans le quatrième exemple, n'est pas déterminée *par un autre agent*. En effet, dans le troisième exemple, c'est la formation qu'a reçue Plum par sa famille et sa communauté qui a déterminé la décision de Plum d'assassiner White. Or, dans le quatrième exemple, aucun autre agent ne semble avoir déterminé la décision de Plum. Cette réponse des compatibilistes est-elle satisfaisante ? Vraisemblablement pas. Premièrement, il serait fortement improbable que, parmi tous les événements du passé qui ont causé la décision de Plum d'assassiner White, aucun agent n'ait jamais été impliqué. Deuxièmement, il ne semble pas que la présence d'un agent, dans les trois premiers exemples, soit essentielle au fait que l'on juge que Plum n'ait pas été libre d'assassiner White (Pereboom 2013, p. 25). Admettons, par exemple, que l'on modifie les trois premiers

exemples, de telle sorte que l'équipe de neuroscientifiques et la famille et la communauté de Plum soient remplacées par une machine sans créateur intelligent ou par un champ de force magnétique<sup>25</sup>. Il ne semble pas que cela change nos intuitions par rapport au fait que l'on juge que Plum n'a pas été libre d'assassiner White.

Comment Pereboom explique-t-il que l'on ne soit pas prêt à dire que Plum ait été libre d'assassiner White dans les quatre exemples présentés ? Pereboom ne croit pas que nos intuitions sont générées par le fait qu'un autre agent ait manipulé Plum. Il croit plutôt que nos intuitions sont générées par le fait que, dans les quatre exemples, la décision de Plum a été déterminée par des facteurs hors du contrôle de Plum: « The salient factor that can plausibly explain why Plum is not responsible in all of the cases is that in each he is causally determined by factors beyond his control to decide as he does ». (Pereboom 2013, p. 25) Ainsi, si Plum ne semble pas libre, ou moralement responsable de l'assassinat de White, puisque Pereboom associe ces deux notions, c'est parce que Plum a été causalement déterminé par des facteurs hors de son contrôle. Réexaminons rapidement chacun des quatre exemples pour s'en convaincre. Dans les deux premiers exemples, nous ne sommes pas prêts à dire que Plum était libre parce qu'une équipe de neuroscientifiques a déterminé la décision de Plum en modifiant ses états mentaux (l'intervention des neuroscientifiques était hors du contrôle de Plum). Dans le troisième exemple, nous ne sommes pas prêts à dire que Plum était libre parce que la famille et la communauté ont déterminé la décision de Plum en le formant d'une certaine façon (Plum n'a eu aucun contrôle sur l'intervention de sa famille et de sa communauté). Finalement, dans le quatrième exemple, nous semblons contraints de dire que Plum n'était pas libre, parce qu'encore ici, la décision de Plum a été déterminée par des facteurs hors du contrôle de Plum et ce, même si aucun agent n'a directement déterminé la décision de Plum. Ainsi, Pereboom soutient que ses cas de manipulation sont menaçants pour le libre arbitre, non pas parce que la décision de l'agent est déterminée par d'autres agents, mais parce que la décision de l'agent est déterminée, plus généralement, par des facteurs hors de son contrôle. Considérant que le premier exemple de Pereboom indique, de manière très intuitive, que Plum n'était pas libre d'assassiner White, les compatibilistes doivent rendre

---

<sup>25</sup> Pereboom fait appel à une machine sans créateur intelligent pour répondre à cette objection compatibiliste (Pereboom 2001). Mele, quant à lui, préfère faire appel à un champ de force magnétique (Mele 2005).

compte de l'absence de libre arbitre de Plum dans le premier exemple de Pereboom et fournir le principe à partir duquel ils différencient le premier du deuxième exemple, le deuxième du troisième exemple, ou le troisième du quatrième exemple. Tels sont les défis que pose l'argument de la manipulation de Pereboom.

### 2.2.1 Deux types de modèles de libre arbitre

Avant d'examiner les réponses que donnent Frankfurt, et Fischer et Ravizza, à l'argument de la manipulation, il est nécessaire de distinguer deux types de modèles de libre arbitre : les modèles historiques et les modèles non historiques<sup>26</sup>. On dira qu'un modèle de libre arbitre est historique lorsqu'il pose au moins une condition historique au libre arbitre. On dira qu'un modèle de libre arbitre est non historique lorsqu'il ne pose aucune condition historique au libre arbitre. Qu'est-ce qu'une condition historique ? Une condition sera dite historique lorsqu'elle porte sur l'histoire de l'agent, sur la façon dont il a acquis ses états mentaux. Le modèle libertarien d'Ekstrom, en exigeant un moment d'indétermination lors de la formation du caractère de l'agent, est un bon exemple de modèle historique. Grâce à sa condition indéterministe, Ekstrom peut discriminer les historiques causaux qui empêchent ou permettent le libre arbitre : les préférences qui ont un historique causal complètement déterminé n'engendrent pas d'actions libres et les préférences qui ont un historique causal indéterminé peuvent engendrer des actions libres. Nous constaterons bientôt, sur la base de nouveaux éléments, que le modèle de Fischer et Ravizza est aussi un modèle historique de libre arbitre. Évidemment, le modèle de Fischer et Ravizza ne s'intéressera pas à l'histoire de l'agent de la même manière qu'Ekstrom, c'est-à-dire en exigeant un indéterminisme dans la formation du caractère de l'agent (ou plus précisément, de ses préférences). Fischer et Ravizza défendent plutôt l'idée qu'un agent, pour être libre, doit avoir pris la responsabilité des mécanismes qui causent ses actions. J'examinerai cette condition sous peu afin d'expliquer en quel sens elle est historique. Les modèles non historiques de libre arbitre, quant à eux, n'accordent aucune importance à la manière dont l'agent a acquis ses états mentaux. Le parfait exemple d'un modèle non historique de libre arbitre est celui de Frankfurt. En effet, celui-ci

---

<sup>26</sup> Cette distinction a été développée, initialement, par Fischer et Ravizza (1998). Fischer et Ravizza distinguent les propriétés historiques (*snapshot properties*) et les propriétés non historiques (*current time-slice properties*). J'ai préféré présenter la distinction en faisant appel à des conditions historiques et à des conditions non historiques.

exige que l'action soit produite par un désir qui lui-même doit être conforme à un désir du second ordre, mais il n'exige rien quant à la manière dont l'agent a acquis le désir du premier ou du second ordre. La distinction entre les modèles historiques et non historiques sera très utile pour distinguer les modèles qui peuvent, ou ne peuvent pas, prétendre rendre compte des cas de manipulation.

### **2.2.2 Les modèles non historiques de libre arbitre : présentation et critique**

Dans cette section, nous verrons que le modèle de Frankfurt ne peut pas relever le défi que pose l'argument de la manipulation, et ce essentiellement parce qu'il s'agit d'un modèle non historique. Un modèle non historique, comme nous venons de le voir, ne s'intéresse pas du tout à la manière dont l'agent a acquis ses états mentaux. Or, les contre-exemples de manipulation misent précisément sur la menace que pose l'intervention d'éléments extérieurs lors de l'acquisition des états mentaux d'un agent. Ainsi, dans les premiers exemples de Pereboom, les neuroscientifiques sont intervenus pour augmenter l'égoïsme de Plum, et c'est cet égoïsme accru qui a engendré la décision d'assassiner White. Sans avoir posé de conditions historiques à son modèle, comment Frankfurt gère-t-il les cas de manipulation ? Frankfurt écrit:

One further condition must be added [to my model]: the person's desires and attitudes have to be relatively well integrated into his general psychic condition. Otherwise they are not genuinely his, but are merely disruptive intruders on his true nature. As long as their interrelations imply that they are unequivocally attributable to him as his desires and attitudes, it makes no difference [...] how he came to have them (Frankfurt 2002, p. 27).

Qu'est-ce que cela veut dire, plus exactement ? Malheureusement, Frankfurt ne précise pas ce qu'il entend par la « condition psychique générale » d'un individu. Cependant, l'idée qu'un agent n'agit pas conformément à sa « véritable nature » lorsque de mauvais génies d'une espèce ou d'une autre interviennent abruptement sur ses états mentaux me semble assez intuitive au premier abord, alors je travaillerai avec elle. Dans le premier exemple de Pereboom, les neuroscientifiques ont augmenté l'égoïsme de Plum seulement quelques minutes avant qu'il décide d'assassiner White. Frankfurt soutiendrait probablement que l'égoïsme qui a engendré la décision d'assassiner White n'était pas vraiment celui de Plum, que cet égoïsme est venu perturber la véritable nature (non égoïste) de celui-ci. Selon Frankfurt, Plum n'était donc pas libre d'assassiner White dans le premier exemple de Pereboom. Dans le deuxième exemple de Pereboom, l'intervention des neuroscientifiques s'est produite du-

rant l'enfance de Plum. Ce dernier a donc vécu toute sa vie avec la programmation des neuroscientifiques. Frankfurt admettrait probablement, ici, que Plum était libre d'avoir assassiné White. En effet, même si l'égoïsme de Plum est le résultat de l'intervention des neuroscientifiques, cet égoïsme fait maintenant partie intégrante de la condition psychique générale de Plum. Depuis l'intervention des neuroscientifiques, Plum est un individu parfois fortement égoïste, et il a agi librement en assassinant White sous l'influence de cet égoïsme (toujours dans le deuxième exemple). Le gain de libre arbitre, selon Frankfurt, se situe ainsi entre le premier et le deuxième exemple de Pereboom.

La nouvelle condition de Frankfurt nous donne-t-elle un principe satisfaisant afin d'expliquer la différence entre un cas de manipulation et un cas de déterminisme normal (tel était le défi lancé par Pereboom) ? Je ne crois pas. À partir de quand un désir ou une attitude devient-il « relativement bien intégré » à la condition psychique générale d'un agent ? La réponse à cette question ne peut pas être arbitraire. Par exemple, il serait insatisfaisant de répondre « après une semaine, on peut dire que le désir de l'agent est relativement bien intégré à sa condition psychique générale ». Comment, alors, Frankfurt justifie-t-il que Plum était libre dans le premier exemple de Pereboom, mais pas dans le deuxième ? Selon Frankfurt, cela a à voir avec les relations qu'entretient l'égoïsme nouvellement acquis de Plum avec le reste de ses désirs et de ses attitudes. À en croire Frankfurt, l'intégration d'un désir ou d'une attitude demande normalement du temps, et Plum n'a pas eu ce temps dans le premier exemple de Pereboom. Acceptons cette idée et examinons une nouvelle version, légèrement modifiée, du premier exemple de Pereboom. Admettons maintenant que l'intervention des neuroscientifiques, en plus d'augmenter l'égoïsme de Plum, fasse en sorte que cet égoïsme soit instantanément intégré à la condition psychique générale de Plum, qu'il fasse si bien partie de sa personnalité que Plum se conçoit lui-même comme une personne égoïste. Qu'en est-il, maintenant ? Nous conservons, même dans ce nouvel exemple, l'intuition que Plum n'était pas libre d'assassiner White. La manipulation des neuroscientifiques se trouve à être un peu plus sophistiquée que dans l'exemple initial, mais nous gardons tout de même l'intuition que Plum n'était pas libre d'avoir assassiné White, car c'est l'égoïsme engendré par les neuroscientifiques qui causé la décision d'assassiner White. Le fait que l'égoïsme soit bien intégré à la condition psychique générale de Plum ne semble rien changer à cela. Par contre, Frankfurt est obligé d'admettre que Plum était libre d'assassiner White, car sa nouvelle condition est remplie dans ce nouvel exemple. Frankfurt ne peut donc plus différencier le premier du deuxième exemple de Pereboom, et Frankfurt se retrouve à court de munitions

pour contrer l'argument de la manipulation de Pereboom et fournir un principe à partir duquel différencier un cas de manipulation d'un cas de déterminisme normal.

En fait, je crois que tous les modèles non historiques de libre arbitre, parce qu'ils ne s'intéressent pas du tout à l'acquisition des états mentaux de l'agent, doivent abandonner le projet de rendre compte des cas de manipulation. En effet, les contre-exemples de manipulation impliquent toujours qu'un agent extérieur soit intervenu lors de l'acquisition d'un état mental, et un modèle non historique de libre arbitre échouera à en rendre compte précisément parce qu'il ne s'intéresse pas à l'histoire causale des états mentaux des agents. Pour Frankfurt, ce sont les désirs du premier et du second ordre qui sont pertinents pour juger qu'une action est libre, mais la critique présentée ci-haut vaut pour n'importe quels états mentaux qui seraient acquis par manipulation, y compris ceux du second ordre qui, selon Frankfurt, garantissent qu'une action est libre. En ne pouvant pas discriminer les cas de manipulation de façon satisfaisante, les modèles non historiques de libre arbitre, comme celui de Frankfurt, échouent à capturer une intuition de base à propos du libre arbitre, soit l'intuition selon laquelle un agent manipulé n'est pas libre. Frankfurt assume complètement le caractère non historique de son modèle. Ainsi, il ne faut pas se surprendre de sa réponse face à l'argument de la manipulation: « It is irrelevant whether [the causes to which we are subject] are operating by virtue of the natural forces that shape our environment or whether they operate through the deliberately manipulative designs of other human agents. We are the sorts of persons we are; and it is what we are, rather than the history of our development, that counts. » (Frankfurt 2002, p. 28). En embrassant le caractère non historique de son modèle, Frankfurt met à jour, je crois, un aspect profondément contre-intuitif de celui-ci. Je crois ainsi qu'il faut rejeter les modèles non historiques de libre arbitre, ceux-ci étant fondamentalement insatisfaisants.

### **2.2.3 Le modèle historique de Fischer et Ravizza: présentation**

Les modèles non historiques, comme celui de Frankfurt, sont des cibles faciles pour l'argument de la manipulation. Heureusement pour les compatibilistes, il est possible de se défendre contre l'argument de la manipulation en adoptant un modèle historique de libre arbitre, un modèle qui s'intéresse à la manière dont l'agent a acquis ses états mentaux. Dans cette section, je présenterai comment Fischer et Ravizza prétendent rendre compte des cas de manipulation en précisant la première condition de leur modèle, de telle sorte qu'elle sera

dorénavant historique<sup>27</sup>. Rappelons les deux conditions que posent Fischer et Ravizza. Selon les deux auteurs, un agent est libre et moralement responsable lorsqu'il fait preuve de contrôle de guidance sur ses actions. Un agent fait preuve de contrôle de guidance sur une action lorsque le mécanisme qui a effectivement produit cette action (1) est le sien et (2) est sensible aux raisons. Lors de la présentation de l'argument de Pereboom, nous avons constaté que la sensibilité aux raisons ne permettait pas de rendre compte du fait que Plum n'était pas libre d'assassiner White. Fischer et Ravizza admettent eux-mêmes que la sensibilité aux raisons ne permet pas de rendre compte de ce type d'exemples de manipulation<sup>28</sup> (Fischer et Ravizza 1998, p. 230). C'est pour cette raison que Fischer et Ravizza développeront la première condition de leur modèle (Fischer et Ravizza 1998, p. 207). Nous verrons que la condition (1), une fois développée, s'intéresse à la manière dont l'agent a acquis ses états mentaux.

Qu'est-ce que cela veut dire, pour un mécanisme, que d'être celui d'un agent ? Jusqu'à présent, nous nous étions contentés d'une compréhension très vague de cette condition. Selon Fischer et Ravizza, un mécanisme devient celui de l'agent lorsqu'il *prend la responsabilité* de ce mécanisme (Fischer et Ravizza 1998, p. 207). Qu'est-ce que cela veut dire que de prendre la responsabilité d'un mécanisme ? Fischer et Ravizza posent trois conditions à ce sujet. Premièrement, un individu doit (A) se voir lui-même comme la source causale de son comportement. Deuxièmement, l'individu doit (B) accepter d'être la juste cible d'attitudes réactives par rapport à son comportement, c'est-à-dire *grosso modo* d'être la cible légitime de blâme et de louange. Troisièmement, l'individu doit (C) respecter les deux premières conditions en se fondant de manière appropriée sur des preuves. Chacune de ces trois conditions

---

<sup>27</sup> Fischer et Ravizza critiquent fortement les modèles non historiques de libre arbitre. Selon eux, le libre arbitre (conféré par le contrôle de guidance) et la responsabilité morale sont des phénomènes authentiquement historiques (Fischer et Ravizza 1998, p. 201).

<sup>28</sup> La sensibilité aux raisons permet de discriminer certains cas de manipulation. Par exemple, admettons qu'un neurochirurgien fasse en sorte que Jérôme devienne frénétique et que la frénésie de Jérôme fasse en sorte qu'il assassine Catherine. Dans ce cas, le mécanisme qui a mené à l'assassinat de Catherine n'était clairement pas sensible aux raisons. Jérôme n'était donc pas libre d'assassiner Catherine. La sensibilité aux raisons posée par Fischer et Ravizza permet ainsi de discriminer les cas de manipulation où un agent devient insensible aux raisons. Fischer et Ravizza ne croient pas que la sensibilité aux raisons soit suffisante pour rendre compte de manière globale des cas de manipulation (Fischer et Ravizza 1998, p. 231). Dans la littérature, rares sont les cas de manipulation où un agent devient complètement insensible aux raisons. Pour cette raison, je mettrai de côté les cas de manipulation que peut gérer la condition (2) du modèle de Fischer et Ravizza.

mériterait de très longues explications tant elles sont complexes. Aux fins de mon argumentation, il me suffira de mettre en évidence deux aspects importants de ses trois conditions.

Premièrement, il faut retenir l'aspect cognitif de la notion de prendre la responsabilité d'un mécanisme, introduit par les conditions (A) et (B).<sup>29</sup> Un individu prend la responsabilité d'un mécanisme lorsqu'il acquiert certaines croyances : (a) des croyances à propos de sa propre efficacité causale dans le monde et (b) des croyances à propos de la légitimité avec laquelle la communauté de l'agent pourra blâmer ou louer ses actions. En réalité, les croyances (a) et (b) ont un contenu assez ciblé. Le modèle de Fischer et Ravizza mise uniquement sur les mécanismes qui produisent l'action. Les croyances (a) et (b) seront aussi orientées vers ces mécanismes. Par exemple, admettons que Jimmy, par désir de vengeance, décide de frapper sa sœur après qu'elle lui ait volé son biscuit. Dans cette situation, on dira que Jimmy a pris la responsabilité du mécanisme psychologique dans lequel s'inscrit son sentiment de vengeance seulement s'il comprend (a) que c'est son mécanisme visant à assouvir sa vengeance qui l'a poussé à frapper sa sœur et (b) que ses parents sont justifiés de le mettre en punition pour avoir agi en fonction de son mécanisme vengeur. Deuxièmement, il faut retenir l'aspect épistémologique de la notion de prendre la responsabilité d'un mécanisme, introduit par la condition (C). La formation des croyances (a) et (b) doit être basée de manière appropriée sur des preuves. Plus précisément, Fischer et Ravizza nous disent que la formation des croyances (a) et (b) doit être basée sur l'expérience de l'individu (Fischer et Ravizza 1998, p. 208). Jimmy, par exemple, pourrait acquérir la croyance (a) grâce à un travail d'introspection, en faisant des liens entre la colère qu'il a ressentie lors du vol de son biscuit et son envie de frapper sa sœur. De plus, Jimmy pourrait acquérir la croyance (b) après avoir été lui-même la cible du mécanisme vengeur de sa sœur.

Le modèle de Fischer et Ravizza, maintenant que la condition (1) est développée, est-il un modèle historique de libre arbitre ? Un modèle de libre arbitre sera dit historique lorsqu'il pose au moins une condition historique. Une condition sera dite historique lorsqu'elle porte sur l'histoire de l'agent, sur la façon dont il a acquis ses états mentaux. Pour prendre la res-

---

<sup>29</sup> Dans le reste du texte, « (A) » et « (B) » feront référence aux deux premières conditions de la notion de prendre la responsabilité d'un mécanisme. « (a) » et « (b) » feront référence aux deux croyances, ou plus exactement aux deux ensembles de croyances, associés à ces conditions. Je caractériserai ces deux croyances de manières plus précises.

ponsabilité d'un mécanisme, l'agent doit remplir les conditions (A), (B) et (C). Est-ce que l'une de ces trois conditions est historique ? Même si elles enrichissent le modèle de Fischer et Ravizza, les conditions (A) et (B) ne semblent pas historiques, ou du moins pas directement. En effet, les conditions (A) et (B) nous disent qu'un agent, pour prendre la responsabilité d'un mécanisme, doit posséder certaines croyances. Prises isolément, les conditions (A) et (B) ne nous disent rien quant à la manière d'acquérir ces dites croyances. Dans l'exemple de Pereboom, les neuroscientifiques pourraient très bien introduire ces deux croyances dans l'esprit de Plum. De ce fait, les conditions (A) et (B) seraient remplies et Plum aurait pris la responsabilité de sa délibération. Si le modèle de Fischer et Ravizza est historique et peut prétendre rendre compte de manière globale des cas de manipulation, c'est essentiellement à cause de la condition (C). Cette condition nous dit que (a) et (b) doivent être acquises à partir de l'expérience de l'individu. Cette condition permet de discriminer les cas de manipulation où un agent extérieur aurait le pouvoir de modifier les états mentaux d'un agent. Dans le cas de Plum, cette condition permet de rendre compte du cas où les neuroscientifiques introduiraient les croyances (a) et (b) dans l'esprit de Plum. Dans un tel cas, les croyances (a) et (b) ne seraient pas fondées sur l'expérience de Plum, et Plum n'aurait donc pas pris la responsabilité de son mécanisme.

Résumons la situation. Le modèle de Fischer et Ravizza nous dit qu'une action est libre (et l'agent moralement responsable) si et seulement si le mécanisme qui a produit l'action (1) est celui de l'agent et (2) est sensible aux raisons. Pour contrer l'argument de la manipulation, Fischer et Ravizza misent sur la condition (1) de leur modèle. Celle-ci est développée de telle sorte que l'on dira qu'un mécanisme est celui de l'agent si et seulement si l'agent a pris la responsabilité de ce mécanisme. Pour prendre la responsabilité d'un mécanisme, un agent doit acquérir certaines croyances. Ces croyances doivent être fondées sur des preuves provenant de l'expérience de l'individu. Fischer et Ravizza sont-ils outillés pour répondre au défi que pose l'argument de la manipulation de Pereboom ? Ce défi, rappelons-le, consiste à identifier le principe à partir duquel il est possible de bloquer le passage d'un cas intuitif de manipulation à un cas normal de déterminisme dans la série d'exemples de Pereboom. La première étape de ce défi consiste, à partir du modèle de Fischer et Ravizza, à rendre compte de l'absence de libre arbitre de Plum dans le premier exemple de Pereboom. La deuxième étape est d'identifier le principe à partir duquel il est possible d'expliquer le « gain » de libre arbitre à travers la série d'exemples. Le modèle de Fischer et Ravizza, comme nous le verrons, échoue à la première étape du défi que lance Pereboom.

#### 2.2.4 Le modèle historique de Fischer et Ravizza : critique

Dans cette section, nous verrons que Fischer et Ravizza échouent à rendre compte de l'absence de libre arbitre de Plum, car la condition (1) ne réussit pas à cerner le lieu où se joue la manipulation dans le cas de Plum, c'est-à-dire son égoïsme. Rappelons-nous le premier exemple de Pereboom. Plum prend la décision d'assassiner White pour des bénéfices personnels et assassine effectivement White suite à sa décision, mais une équipe de neuroscientifiques a été capable de manipuler les états mentaux de Plum à l'aide de technologies radio. Avant que Plum ne commence à délibérer à propos de l'assassinat de White, les neuroscientifiques ont appuyé sur un bouton, mettant Plum dans un état neuronal faisant en sorte que son prochain raisonnement serait fortement égoïste. Si les neuroscientifiques n'étaient pas intervenus, Plum n'aurait pas assassiné White. Plum est ainsi victime de manipulation et, de ce fait, ne semble pas avoir été libre d'assassiner White. Est-ce que le modèle de Fischer et Ravizza peut en rendre compte ? En d'autres mots, est-ce que Fischer et Ravizza peuvent nous montrer que le mécanisme délibératif qui a produit la décision d'assassiner White *n'est pas* (1) celui de Plum ou *n'est pas* (2) sensible aux raisons ? Nous avons déjà déterminé que la délibération de Plum était bel et bien sensible aux raisons. Dans un monde possible où Plum aurait eu des raisons de ne pas assassiner White, Plum aurait reconnu ces raisons de ne pas assassiner White, aurait décidé de ne pas assassiner White et n'aurait pas assassiné White. Ainsi, l'égoïsme nouvellement acquis de Plum ne change rien à la sensibilité aux raisons de la délibération de Plum.

La délibération de Plum était-elle celle de Plum ? Malgré l'analyse développée de Fischer et Ravizza, et malgré la mise en place des conditions (A), (B) et (C), il semble que ce soit bel et bien le cas. En effet, la manipulation des neuroscientifiques se joue uniquement au niveau de l'égoïsme de Plum, et pas au niveau du mécanisme délibératif. Or, le modèle de Fischer et Ravizza s'intéresse uniquement aux mécanismes ayant produit l'action. L'introduction de la condition historique (C) n'aidera pas Fischer et Ravizza, parce que la condition (C) s'intéresse uniquement au mécanisme délibératif de Plum. Plus précisément, la condition (C) porte sur les croyances (a) et (b), qui elles-mêmes portent sur le mécanisme délibératif de Plum. En fait, toutes les conditions posées par Fischer et Ravizza sont orientées vers les mécanismes ayant produit l'action. Le modèle de Fischer et Ravizza, en s'intéressant uniquement aux mécanismes ayant produit l'action, délaisse les facteurs « périphériques » aux mécanismes. Des exemples de facteurs périphériques seraient l'humeur de l'agent, ses émo-

tions, son niveau de stress et, dans le cas qui nous intéresse, son égoïsme. Parce qu'il mise sur autre chose que le mécanisme de délibération ayant produit l'action, le cas de manipulation de Pereboom échappe donc au modèle de Fischer et Ravizza. Examinons à nouveau l'exemple de Plum pour s'en convaincre.

Dans le premier exemple de Pereboom, les neuroscientifiques sont intervenus seulement quelques minutes avant que Plum ne commence à délibérer à propos de l'assassinat de White. Plum a donc eu plusieurs années, avant l'intervention des neuroscientifiques, pour prendre la responsabilité de son mécanisme délibératif. Par exemple, il est raisonnable d'affirmer que Plum a acquis, via ses propres expériences (C), les croyances (a) que son mécanisme délibératif cause certaines de ses actions et (b) qu'il est légitime de blâmer ou louer les actions qui résultent de son mécanisme délibératif. La condition (1) du modèle de Fischer et Ravizza est donc remplie. En d'autres mots, l'augmentation de l'égoïsme de Plum ne semble rien changer au fait que Plum ait pris la responsabilité de son mécanisme délibératif. En ajoutant les croyances (a) et (b) dans l'exemple de Pereboom, il est très plausible que Plum se sente horrifié d'avoir assassiné White. D'un point de vue extérieur, personne ne pourrait se douter que Plum a été manipulé. Pis encore, Plum pourrait se livrer lui-même à la police sitôt après avoir commis cet acte qu'il juge horrible. Est-ce problématique ? Je ne crois pas. La manipulation des neuroscientifiques n'exclut pas la culpabilité de Plum. En somme, selon le modèle de Fischer et Ravizza, Plum était libre et moralement responsable de l'assassinat de White. En ne pouvant pas rendre compte de l'absence de libre arbitre dans le premier cas de Pereboom, Fischer et Ravizza échouent à relever le défi que pose l'argument de la manipulation.

Fischer et Ravizza n'ont pas dit leur dernier mot. En fait, il faut savoir que l'argument de la manipulation de Pereboom (2001) a été publié trois ans après la publication du livre de Fischer et Ravizza (1998). Fischer et Ravizza n'ont donc pas conçu leur modèle pour répondre spécifiquement à cet argument. Dans un article subséquent, Fischer (sans Ravizza) répond à Pereboom :

Pereboom basically asks the compatibilist to point to the place (after Case 1) along the slippery slope where responsibility emerges. My answer: there is no such place, as Pereboom suggests. Rather, on a plausible understanding of the case, Professor Plum is morally responsible in Case 1. Thus, there is no impediment to saying that Plum is responsible in Case 4 (and, in general, in the context of causal determinism). (Fischer 2004, p. 156)

Fischer admet, en un sens, ne pas pouvoir relever le défi de Pereboom. Pour être plus exact, selon lui, il n'y a pas de défi à relever. Au lieu de rendre compte de l'absence de libre arbitre dans le premier exemple de Plum, Fischer va tenter de nous convaincre que Plum était bel et bien libre, et moralement responsable, d'assassiner White. C'est pour le moins audacieux. Fischer reconnaît, semble-t-il, l'aspect fortement contre-intuitif de sa position. En effet, pour se sortir d'embarras, Fischer va poser une distinction entre *la responsabilité morale de ses actions* (conférée par le contrôle de guidance) et la propriété d'*être moralement blâmable pour ses actions* (Fischer 2004, p. 157). L'idée générale, derrière cette distinction, c'est que la responsabilité morale n'implique pas qu'un agent est digne de blâme (ou de louange) pour ses actions. Fischer dira que la responsabilité morale n'est qu'une « passerelle » pour blâmer ou louer les actions d'un agent. Qu'est-ce que cela veut dire ? Que faut-il de plus pour être moralement blâmable de ses actions ? À cette dernière question, Fischer répond :

In my view, further conditions need to be added to mere guidance control to get to blameworthiness; these conditions may have to do with the circumstances under which one's values, beliefs, desires, and dispositions were created and are sustained, one's physical and economic status, and so forth. Professor Plum, it seems to me, is not blameworthy, even though he is morally responsible. (Fischer 2004, p. 158)

Fischer soutient ainsi que Plum n'est pas moralement blâmable de l'assassinat de White dans le premier exemple de Pereboom, et ce même s'il est libre d'accomplir cette action et moralement responsable de l'avoir fait. À l'aide de sa distinction, Fischer prétend pouvoir expliquer pourquoi Plum n'est pas moralement blâmable de l'assassinat de White dans le premier exemple de Pereboom, et pourquoi Plum devient moralement blâmable de l'assassinat de White dans le quatrième exemple de Pereboom : « Note that there is no difference with respect to the minimal control conditions for moral responsibility in Cases 1 through 4 – the threshold is achieved in all the cases. But there are (or may be, for all that has been said in Pereboom's descriptions) wide disparities in the conditions for blameworthiness. » (Fischer 2004, p. 158) Fischer n'élabore pas à propos des conditions qui expliqueraient pourquoi Plum devient moralement blâmable dans le quatrième exemple de Pereboom.

Même en faisant abstraction du fait que Fischer ne définit pas les conditions supplémentaires qu'il faut remplir pour être moralement blâmable de ses actions, la solution qu'il avance est déjà hautement problématique. Fischer met de l'avant une notion de responsabilité morale qui est considérablement différente de celle généralement admise. En effet, la no-

tion standard de responsabilité morale est définie précisément en termes de blâme et de louange, ainsi qu'il ressort de l'entrée « Moral Responsibility » de la *Stanford Encyclopedia of Philosophy*: « to be morally responsible for something, say an action, is to be worthy of a particular kind of reaction—praise, blame, or something akin to these—for having performed it » (Eshleman 2014). En distinguant la responsabilité morale de la propriété d'être blâmable de ses actions, Fischer s'éloigne drastiquement de la notion standard de responsabilité morale, celle dont il était question dans le débat sur le libre arbitre<sup>30</sup>. Pour cette raison, je considère que la réponse de Fischer est complètement *ad hoc* et ne permet pas de le sortir de la position embarrassante dans laquelle il se retrouve lorsqu'il affirme que Plum est libre et moralement responsable de l'assassinat de White dans le premier exemple de Pereboom. Parce qu'il échoue à rendre compte du cas de manipulation de Pereboom à moins d'avoir recours à une solution qui exige une révision drastique de la notion de responsabilité morale, je crois que le modèle de libre arbitre de Fischer et Ravizza n'est pas satisfaisant.

### 2.2.5 Une vue d'ensemble

Avant de conclure, j'aimerais prendre un peu de recul et examiner la forme générale que prend habituellement l'argument de la manipulation. Cela permettra de mettre en évidence les problèmes auxquels sont confrontés Frankfurt, Fischer et Ravizza. La formulation de Pereboom (2013) est devenue très populaire, mais elle a le défaut, à mon avis, de nous faire perdre la problématique de vue. Les arguments de la manipulation sous-tendent tous un raisonnement de la forme suivante :

- (1) Les agents qui sont manipulés ne sont pas libres.
- (2) Il n'existe aucune différence fondamentale entre un agent manipulé et un agent qui est déterminé.
- (3) Donc, les agents qui sont déterminés ne sont pas libres.

La prémisse (1) est basée sur l'intuition qu'un individu qui est sous l'emprise d'un manipulateur ne peut pas être qualifié légitimement de « libre ». Dans l'argument de Pereboom, la prémisse (1) se manifeste à travers le premier exemple de Plum. La prémisse (2), quant à

---

<sup>30</sup> Pour une critique plus élaborée de la distinction que pose Fischer, voir Pereboom (2007a) dans Fischer et al. (2007). La réponse de Fischer (2007) se trouve aussi dans Fischer et al. (2007), tout comme la critique de cette réponse (Pereboom 2007b).

elle, est basée sur l'intuition qu'il n'existe aucune différence fondamentale entre une action qui est déterminée *par un autre agent* et une action qui est déterminée *par les événements du passé en conjonction avec les lois de la nature*. Dans l'argument de Pereboom, la prémisse (2) se manifeste à travers le deuxième, le troisième et le quatrième exemple de Plum. Comment réagissent Frankfurt, Fischer et Ravizza à cet argument? Frankfurt accepte la prémisse (2): « It is irrelevant whether [the causes to which we are subject] are operating by virtue of the natural forces that shape our environment or whether they operate through the deliberately manipulative designs of other human agents. We are the sorts of persons we are; and it is what we are, rather than the history of our development, that counts. » (Frankfurt 2002, p. 28) Frankfurt se voit ainsi contraint de nier la prémisse (1) pour échapper à l'argument de la manipulation. Je crois, pour ma part, qu'abandonner la prémisse (1) est un prix trop lourd à payer, et j'ai préféré ainsi rejeter le modèle de Frankfurt. Fischer et Ravizza, quant à eux, sont beaucoup plus réfractaires que Frankfurt à abandonner la prémisse (1). Fischer et Ravizza, en nous disant qu'un mécanisme devient celui de l'agent lorsqu'il a pris la responsabilité de ce mécanisme, essaient de démontrer la fausseté de la prémisse (2). J'ai soutenu que la solution de Fischer et Ravizza n'était pas satisfaisante parce qu'elle exigeait une révision trop importante de la notion de responsabilité morale. Ce faisant, Fischer et Ravizza ne réussissent pas, à mon avis, à répondre à l'argument de la manipulation de façon adéquate.

Bien que les modèles de Frankfurt et de Fischer et Ravizza ne réussissent pas à répondre de façon satisfaisante à l'argument de la manipulation, pourquoi serait-ce nécessairement le cas pour tous les modèles compatibilistes de libre arbitre ? Je n'ai aucune réponse à offrir à cette question. Les modèles de Frankfurt et de Fischer et Ravizza sont des modèles compatibilistes très respectés, mais on ne peut pas inférer, de leur échec, que le compatibilisme est condamné à se buter aux contre-exemples de manipulation pour l'éternité. Même s'il n'existe, à ma connaissance, aucun modèle qui échappe à l'argument de la manipulation, je ne peux qu'admettre, ici, les limites de ma critique.

## **2.3 Conclusion**

L'objectif de ce chapitre était d'exposer une critique convaincante du compatibilisme, la thèse selon laquelle le libre arbitre est compatible avec la vérité du déterminisme. J'ai commencé par présenter l'argument de la conséquence de van Inwagen (1986). Selon cet

argument, si le déterminisme est vrai, alors notre manque de contrôle sur les événements du passé et sur les lois naturelles engendre un manque de contrôle sur nos actions. Cet argument est fondé sur cinq éléments : la vérité du déterminisme, la fixité du passé, la fixité des lois de la nature, la validité de la règle ( $\alpha$ ) et la validité de la règle ( $\beta$ ). Comme nous l'avons vu, seule la validité de la règle ( $\beta$ ) est controversée. Après avoir examiné l'objection de McKay et Johnson (1996), j'ai soutenu que la règle ( $\beta$ ) était bel et bien valide et que l'argument de la conséquence de van Inwagen était fondé. Ainsi, le libre arbitre ne semble pas compatible avec le déterminisme, malgré les prétentions des compatibilistes. Par la suite, j'ai présenté l'argument de la manipulation de Pereboom (2013). Cet argument, construit à partir de quatre exemples, lance deux défis aux compatibilistes : (1) rendre compte de l'absence de libre arbitre de Plum dans le premier exemple de manipulation et (2) fournir le principe à partir duquel ils différencient les cas de manipulation des cas de déterminisme « ordinaire ». Après avoir posé une distinction entre les modèles historiques et les modèles non historiques de libre arbitre, et après avoir examiné les réponses de Frankfurt et de Fischer et Ravizza à l'argument de la manipulation, j'ai soutenu que les compatibilistes échouaient à relever le défi lancé par Pereboom. Sur la base de ces deux arguments, j'ai jugé qu'il valait mieux rejeter les modèles compatibilistes de libre arbitre. Ma critique, évidemment, n'est pas complète. Les modèles compatibilistes sont très variés, et une critique tout à fait charitable nécessiterait un examen minutieux de l'ensemble des modèles compatibilistes. En ayant examiné et critiqué deux modèles parmi les plus respectés de la littérature compatibiliste, je crois tout de même avoir donné de bonnes raisons de rejeter le compatibilisme.

Dans le prochain chapitre, je présenterai une critique des modèles libertariens de libre arbitre. Selon le libertarianisme, le libre arbitre existe même s'il est incompatible avec le déterminisme. Je soutiendrai, grâce à l'argument de la chance, que le libre arbitre n'est pas le bienvenu dans un monde indéterministe. Par la suite, je soutiendrai la vérité du quasi-déterminisme, la thèse selon laquelle nous sommes entièrement déterminés au niveau de notre existence. Sur la base de ces deux arguments, je soutiendrai que le libre arbitre n'est pas plus compatible avec l'indéterminisme qu'il ne l'est avec le déterminisme. La voie sera ainsi ouverte pour la défense d'une thèse sceptique, la thèse illusionniste de Smilansky (2000), que j'exposerai au chapitre 4.

## Chapitre 3 : Contre le libertarianisme

### 3.0 Introduction

Dans le chapitre précédent, nous avons rejeté les modèles compatibilistes de libre arbitre parce que le libre arbitre n'est pas, arguments à l'appui, compatible avec le déterminisme, et parce que les modèles compatibilistes examinés, pourtant parmi les plus convaincants dans la littérature, ne réussissent pas à rendre compte de manière adéquate des cas de manipulation. Le rejet du compatibilisme nous conduit naturellement vers le libertarianisme, la thèse selon laquelle le libre arbitre existe même s'il est incompatible avec le déterminisme. Le libertarianisme, peu importe comment il est défendu, est confronté à deux problèmes importants. Premièrement, les libertarianistes doivent nous expliquer comment le libre arbitre est possible dans un monde indéterminé. Si nos actions sont indéterminées, comment peut-on soutenir qu'elles sont libres ? Deuxièmement, les libertarianistes doivent nous donner des raisons de croire en la fausseté du déterminisme. En effet, selon le libertarianisme, le libre arbitre peut exister uniquement si le déterminisme est faux. Or, le déterminisme est une thèse largement acceptée dans le milieu scientifique. Dans ce chapitre, je soutiendrai que les libertarianistes échouent à relever ces deux défis. Ce chapitre se donne ainsi comme objectif de proposer une critique convaincante du libertarianisme. Pour réaliser cet objectif, je procéderai en deux temps. Premièrement, j'exposerai l'argument de la chance. Dans cette section, mon objectif sera d'évaluer si le libertarianisme est cohérent en exigeant de l'indéterminisme dans la chaîne causale menant aux actions libres. Après avoir examiné les failles de cet argument, je me tournerai vers une critique qui ne fait pas appel à la notion de chance, critique selon laquelle il n'est pas possible de structurer nous-mêmes notre caractère, contrairement à ce que prétendent certains libertarianistes. Deuxièmement, je défendrai la vérité du quasi-déterminisme, la thèse selon laquelle nous sommes entièrement déterminés *au niveau de notre existence*. L'objectif sera de mettre en évidence d'une part que le quasi-déterminisme est vrai, ce qui rend le libertarianisme invalide, et d'autre part que le libre arbitre est incompatible avec le quasi-déterminisme.

### 3.1 L'argument de la chance

Dans cette section, j'examinerai si la notion libertarianiste de libre arbitre est cohérente. Pour ce faire, je ferai appel à l'argument de la chance. L'idée générale derrière l'argu-

ment de la chance ressemble à ceci : les libertarianistes, en exigeant de l'indéterminisme dans la chaîne causale menant aux actions libres, introduisent de la chance dans cette même chaîne causale, de telle sorte qu'il n'est plus possible de considérer qu'une action est libre. Pour contrer ce problème, certains philosophes décident de ne pas placer l'indéterminisme directement dans la chaîne causale menant à l'action, mais plutôt dans la chaîne causale menant à la formation du caractère des agents<sup>31</sup> (Kane 1996; Ekstrom 2000; Mele 1995). Cette stratégie suffit-elle ? Pour répondre à cette question, j'exposerai premièrement l'argument de la chance tel que présenté par van Inwagen (2000). Deuxièmement, j'examinerai comment Ekstrom, en distinguant trois sens au mot « chance » répond à l'argument de la chance (Ekstrom 2011). Troisièmement, je soutiendrai que le modèle d'Ekstrom, même s'il résiste plutôt bien à l'argument de la chance, est invalide parce qu'il exige que nous puissions structurer nous-mêmes notre caractère, ce qui est impossible. Pour ce faire, je présenterai un argument inspiré de Galen Strawson (2013).

Il existe plusieurs façons de présenter l'argument de la chance. J'ai choisi d'utiliser l'argument de van Inwagen parce qu'il est possible de le formuler de manière plus intuitive et parce qu'Ekstrom répond explicitement à l'argument de la chance tel qu'avancé par van Inwagen. Le débat n'en sera que plus intéressant. L'objectif, ici, est d'évaluer si le libre arbitre libertarianiste est cohérent parce que l'exigence de départ des libertarianistes, la présence d'un indéterminisme dans la chaîne causale menant aux actions libres, implique de la chance qui semble rendre impossible toute action libre. L'argument peut se formuler ainsi :

- (1) Pour les libertarianistes, si les actions d'un agent ne sont pas indéterminées, alors l'agent ne peut pas être libre.
- (2) Cependant, si les actions d'un agent sont indéterminées, alors la façon dont l'agent agit est une question de chance.
- (3) Si la façon dont l'agent agit est une question de chance, alors on peut difficilement dire que l'agent est libre.
- (4) En conséquence, si les actions d'un agent sont indéterminées, alors on peut difficilement dire que l'agent est libre.

---

<sup>31</sup> Ces auteurs ajoutent que le caractère, formé de manière indéterminée, cause de manière déterminée les actions des agents. Il s'agit d'un modèle de libre arbitre à deux étapes, qui fait appel à la fois à l'indéterminisme et au déterminisme. Ce type de modèle est de plus en plus populaire, d'où mon choix d'exemplifier le libertarianisme à l'aide du modèle d'Ekstrom.

(5) Donc, les modèles indéterministes de libre arbitre impliquent qu'il n'existe aucune action libre.

Comment van Inwagen défend-il son argument ? Contrairement à ses habitudes, van Inwagen préfère ici présenter son argument de façon intuitive<sup>32</sup>, sous la forme d'une histoire. Alice fait face à un choix difficile entre mentir et dire la vérité. Elle décide, après réflexion, de dire la vérité. Présumons maintenant que son choix était libre. Ainsi, (1) selon les libertariens, le choix d'Alice était indéterminé. Après qu'Alice ait décidé de dire la vérité, Dieu intervient et rembobine le temps jusqu'au moment où Alice délibérait pour savoir si elle allait mentir ou dire la vérité. Dieu laisse ensuite s'écouler le temps normalement, jusqu'à ce qu'Alice en soit rendue à prendre sa décision. Décidera-t-elle de mentir ou de dire la vérité, cette fois ? Considérant que la première décision d'Alice était indéterminée, la seconde le sera aussi. Ainsi, il est impossible de savoir si Alice décidera de mentir ou de dire la vérité. Tout ce qu'on peut dire, c'est qu'elle pourrait choisir de dire la vérité et qu'elle pourrait choisir de mentir. Cette fois, Alice décide de mentir. Admettons maintenant que Dieu refasse cette petite expérience 1000 fois. Ultimement, nous apprenons qu'Alice a décidé de mentir 500 fois et a décidé de dire la vérité 500 fois. Que pouvons-nous apprendre de ces résultats ? Certainement, nous pouvons conclure que la décision d'Alice n'est pas déterminée par l'état initial d'Alice. Tout ce qu'il nous est permis de dire, c'est qu'il semble y avoir une probabilité de départ (*ground-floor*) de 50% qu'Alice décide de mentir et une probabilité de départ de 50% qu'elle décide plutôt de dire la vérité. Que se passera-t-il lorsque Dieu renversera le temps une autre fois ? Impossible de le savoir. Nous sommes ainsi frappés par l'impression que (2) la décision d'Alice sera aléatoire, que ce sera une question de chance. Or, si la décision d'Alice est une question de chance dans cette mille et unième reprise temporelle, la décision d'Alice était aussi une question de chance dans la situation originale. En effet, que l'on ne connaisse pas, à ce moment, la probabilité de départ qu'Alice décide de mentir ou de dire la vérité ne change rien au fait qu'il s'agissait aussi, à ce moment, d'une question de chance. Mais alors, comment peut-on affirmer qu'Alice était libre de mentir ou de dire la vérité si le résultat de sa délibération était une question de chance ? Selon van Inwagen, (3) il n'est pas possible d'affirmer que la décision d'Alice était libre, précisément parce que sa décision était une question de chance. (4) Il

---

<sup>32</sup> Van Inwagen (1986) avait déjà présenté son argument de la chance de manière très formelle. Dans l'article qui nous intéresse (van Inwagen 2000), il ne croit plus que son argument formel soit valide. Il préfère donc présenter son argument de la chance sous une forme plus intuitive, et selon lui, plus convaincante.

semble ainsi difficile de dire qu'Alice était libre de dire la vérité si sa décision était indéterminée. Comme (1) affirme qu'on ne peut pas dire non plus qu'Alice était libre si sa décision n'était pas indéterminée, on doit conclure (5) qu'il n'y a pas d'actions libres. Examinons comment Ekstrom répond à cet argument.

### 3.1.1 L'objection d'Ekstrom à l'argument de la chance

Ekstrom, étant libertarieniste, ne souhaite pas que son modèle implique l'inexistence du libre arbitre. Pour attaquer l'argument de la chance et bloquer la conclusion (5), Ekstrom doit s'attaquer à la vérité de l'une des prémisses ou à la validité de l'argument. Examinons comment elle s'y prend. Ekstrom est évidemment d'accord avec la prémisse (1), celle-ci étant le point de départ de son investigation sur le libre arbitre. La prémisse (4), quant à elle, découle des prémisses (2) et (3). Les prémisses (2) et (3), même si elles paraissent intuitives, sont controversées. Ekstrom décide de ne pas s'attaquer directement à la vérité de ces deux prémisses<sup>33</sup>. Elle attaquera plutôt la validité de l'argument. Selon elle, l'argument de van Inwagen n'est pas valide, car van Inwagen commet un « sophisme de la synonymie »<sup>34</sup>. En d'autres mots, Ekstrom accuse van Inwagen de ne pas utiliser la même notion de *chance* dans les prémisses (2) et (3), ce qui fait en sorte que la conclusion ne découle pas logiquement des prémisses. Pour le démontrer, Ekstrom distinguera trois notions de chance. Chacune de ces notions permet uniquement à l'une ou l'autre des prémisses d'être vraies, mais aucune de ces notions ne réussit à rendre simultanément vraies les deux prémisses. Examinons l'analyse d'Ekstrom plus en détail.

La première notion de chance qu'Ekstrom soulève présente la chance comme une force indépendante, un déterminateur sans but (*purposeless determiner*) qui rend compte d'événements qui seraient, sans cela, laissés sans explication (Ekstrom 2011, p. 375). Cette notion de chance, ajoute Ekstrom, semble traiter la chance comme un mystérieux agent lâché dans le monde et qui possède sa propre force causale. Cette notion de chance est pré-

---

<sup>33</sup> Indirectement, par contre, Ekstrom attaque bel et bien la vérité de chacune de ces prémisses. Plusieurs éléments, dans sa défense contre l'argument de la chance, ont déjà été exposés dans le premier chapitre. Je les exposerai plus brièvement ici.

<sup>34</sup> Ekstrom n'utilise pas le terme « sophisme de la synonymie », mais je trouve l'expression particulièrement appropriée.

sente dans des énoncés comme « Le résultat des élections est maintenant entre les mains de la chance » ou encore « Laissons la chance décider qui elle choisira d'aimer ». Ekstrom soulève, à juste titre, qu'aucun libertarien ne croit en l'existence d'une telle « chance ». Aux fins de l'argument, admettons qu'il existe un tel déterminateur sans but. Dans ce cas, la prémisse (3) semble vraie, car un agent qui agit sous l'influence d'un déterminateur sans but ne semble pas libre, ou du moins, très difficilement. Par contre, la prémisse (2) est fautive. En effet, l'indéterminisme, c'est-à-dire la thèse selon laquelle une action n'est pas déterminée par les événements du passé en conjonction avec les lois de la nature, n'implique pas qu'une mystérieuse force causale ait produit cette action. Ainsi, il ne semble pas que l'argument de van Inwagen soit valide s'il est fondé sur une telle notion de chance.

La deuxième notion de chance qu'Ekstrom soulève est probabiliste. Un événement qui est soumis à la chance, en ce sens, est un événement qui a une probabilité de se produire entre 0 et 1. Supposons, aux fins de l'argument, que nos actions sont gouvernées par des lois probabilistes<sup>35</sup>. Ainsi, la prémisse (2) de l'argument de van Inwagen semble vraie. En effet, il est admis que l'indéterminisme implique un certain degré de chance au sens où certains événements ont des probabilités entre 0 et 1 de se produire. Par contre, la prémisse (3) est fautive selon Ekstrom. Le modèle de libre arbitre d'Ekstrom est développé précisément en ce sens, c'est-à-dire en tentant de démontrer que la présence de chance (au sens probabiliste) n'implique pas l'inexistence du libre arbitre. Ekstrom nous rappelle que sa notion de libre arbitre implique seulement une certaine forme de contrôle par le soi (*self*) et l'existence de possibilités alternatives. Dans le modèle d'Ekstrom, le contrôle par le soi est généré par le fait que le caractère détermine de manière nécessaire la façon dont l'agent agit. Les possibilités alternatives, quant à elles, sont localisées lors de la formation du caractère de l'agent. Ekstrom conclut: « this is a harmless sense of "chance" in regard to our self-governance of our decisions, as it does not make those decisions purposeless, unguided, or haphazard » (Ekstrom 2011, p. 376). Ainsi, il ne semble pas que l'argument de van Inwagen soit valide s'il est fondé sur une telle notion de chance non plus.

---

<sup>35</sup> Ekstrom précise qu'elle ne croit pas que toutes les actions soient gouvernées par des lois, mais elle ne le justifie pas dans son article. De toute façon, cela nous éloignerait de son argument.

La troisième et dernière notion de chance qu'Ekstrom soulève nous dit qu'un événement soumis à la chance est un événement sans but (*purposeless*), ne faisant pas partie du plan de l'agent et ne pouvant pas recevoir de justifications rationnelles. Selon cette notion de chance, la prémisse (3) semble vraie. En effet, on voit mal comment une action peut être libre si elle n'est pas dirigée vers une fin de l'agent et si elle ne peut pas être justifiée par des raisons. Par contre, en ce sens, la prémisse (2) est fautive, car l'indéterminisme n'implique pas qu'une action ne puisse pas être justifiée par des raisons. Le modèle d'Ekstrom le démontre bien. Dans le cas de Sophie, par exemple, qui délibère pour savoir si elle va donner 800\$ à une œuvre de charité ou plutôt partir en voyage au Mexique, le résultat de sa décision est indéterminé. Malgré cela, il lui est possible de donner des raisons pour chacune des décisions qu'elle prendra. Si Sophie décide de donner 800\$ à une œuvre de charité, elle dira que c'est parce qu'il lui importe d'être généreuse et qu'elle désire aider les plus démunis, et si Sophie décide de partir au Mexique, elle dira que c'est parce qu'elle vit beaucoup de stress et qu'elle a besoin de se reposer sur la plage. Encore ici, il ne semble pas que l'argument de van Inwagen puisse être valide s'il est fondé sur cette notion particulière de chance.

### **3.2 Est-il possible de structurer son caractère ?**

Ekstrom, en distinguant trois significations différentes du mot « chance », démontre que l'argument de van Inwagen n'est pas valide et réussit à mettre le fardeau de la preuve sur les épaules des sceptiques. Ces derniers, pour nous convaincre de l'inexistence du libre arbitre à l'aide d'un argument de la chance, doivent fournir une analyse détaillée de la notion de chance qu'ils privilégient. En attendant la mise en place d'une telle notion de chance, il semble que l'argument de la chance soit voué à l'échec. Ainsi, peut-être est-il faux d'affirmer que l'indéterminisme, dans le modèle d'Ekstrom, introduit de la chance au point où il n'est plus possible de considérer qu'un agent agit librement. Par contre, même si l'indéterminisme n'introduit pas une forme menaçante de chance dans le modèle d'Ekstrom, il est possible d'attaquer ce modèle sur un autre front, c'est-à-dire sur l'idée selon laquelle nous structurons nous-mêmes notre caractère. Dans cette section, j'évaluerai cette idée et soutiendrai que nous n'avons aucune bonne raison de croire que c'est effectivement le cas.

Réexaminons rapidement le modèle d'Ekstrom. Ekstrom localise l'indéterminisme durant les délibérations menant à la formation de nos préférences, ces dernières étant une partie essentielle de notre caractère. Ekstrom prétend échapper au problème du contrôle, car

durant la délibération, qui est un processus indéterministe, notre faculté évaluative entre en jeu. Cette faculté, étant aussi une partie essentielle du soi, permet à Ekstrom d'affirmer que nous pouvons structurer nous-mêmes notre caractère. En quoi cela aide-t-il Ekstrom à répondre au problème du contrôle ? Si l'indéterminisme semble menaçant pour le libre arbitre, c'est essentiellement parce qu'il laisse un certain vide explicatif, ce qui laisse croire que les actions se produisent aléatoirement ou accidentellement : « Incompatibilist [...] must examine the notion of control and explain how, on their own particular account, the free agent's act is grounded in his agency and is not a haphazard or accidental happening. » (Ekstrom 2000, p. 86) Si nos délibérations ne sont pas déterminées, comment expliquer qu'elles aient causé une décision, et comment expliquer qu'elles aient causé telle décision et pas une autre ? Si l'on en croit Ekstrom, c'est l'agent lui-même qui comble le vide laissé par l'indéterminisme, en se servant de sa faculté évaluative pour orienter (sans déterminer) le résultat des délibérations portant sur son caractère. Examinons cette idée plus en détail.

Pour Ekstrom, s'il est possible de structurer soi-même son caractère, c'est parce que le soi participe à toutes les étapes de la chaîne causale menant à l'action. Réexaminons brièvement cette fameuse chaîne causale. Celle-ci est constituée de quatre maillons : (1) la délibération, (2) les préférences et les adhésions<sup>36</sup>, (3) l'intention et (4) l'action. Ekstrom définit le soi comme comprenant à la fois notre faculté évaluative et notre caractère (Ekstrom 2000, p. 122, 123). La faculté évaluative est la source de l'évaluation des raisons ainsi que des décisions qui s'ensuivent (p. 115). Ekstrom en parle aussi comme de l'intellect (p. 110). Le caractère est la somme de nos préférences et de nos adhésions (p. 114). Est-ce que le soi, au sens d'Ekstrom, participe à la formation de chacun des maillons de la chaîne causale ? Il semble que oui. D'une part, la délibération est soumise à la faculté évaluative, et donc au soi. D'autre part, les préférences et les adhésions sont elles-mêmes le résultat des délibérations, et donc aussi soumises au soi. On voit réapparaître, ici, l'idée selon laquelle *les agents structurent eux-mêmes leur caractère*, une idée sur laquelle Ekstrom insiste très souvent. Finalement, les liens causaux entre (2) et (3) et entre (3) et (4) sont déterministes. De ce fait, Ekstrom s'assure une participation du soi dans (3) et dans (4), en ce sens que c'est le caractère qui détermine de manière nécessaire (3) et (4).

---

<sup>36</sup> Le concept d'adhésion n'a pas été discuté jusqu'à présent. Il s'agit, *grosso modo*, d'une forme plus ferme de croyance, de la même manière que les préférences sont une forme plus nette de désir.

### 3.2.1 L'argument de base

Le modèle d'Ekstrom met donc de l'avant une idée bien précise : si nous sommes libres, c'est parce que nous pouvons structurer notre caractère. Mais pouvons-nous vraiment structurer notre caractère ? Je soutiendrai que non. Pour ce faire, je soumettrai un argument inspiré de « l'argument de base » (*basic argument*) de Strawson (2013, p. 42)<sup>37</sup>. Je commencerai par présenter l'argument de Strawson, pour ensuite le confronter au modèle d'Ekstrom.

- (1) En nous intéressant aux actions libres, nous sommes intéressés par les actions qui sont accomplies pour une raison.
- (2) Quand quelqu'un agit pour une raison, il agit en fonction de ce qu'il est, psychologiquement parlant.
- (3) Donc, si l'on veut dire d'une personne qu'elle agit librement, elle doit avoir librement choisi ce qu'elle est, psychologiquement parlant, à certains égards.
- (4) Mais pour choisir librement ce qu'elle est, psychologiquement parlant, elle doit avoir structuré elle-même ce qu'elle est, psychologiquement parlant, au moins à certains égards.
- (5) Or, aucune personne ne peut structurer elle-même ce qu'elle est, psychologiquement parlant, et ce à aucun égard, à moins d'être déjà pourvue de certains principes de choix (P1) – des préférences, des valeurs, des pro-attitudes, des idéaux – sur la base desquels elle a formé ce qu'elle est, psychologiquement parlant.
- (6) Mais alors, si l'on veut dire d'une personne qu'elle structure elle-même ce qu'elle est, psychologiquement parlant, elle doit avoir choisi les principes de choix (P1) sur la base desquels elle a formé ce qu'elle est, psychologiquement parlant.
- (7) Pour cela, cette personne doit déjà être pourvue de principes de choix (P2), sur la base desquels elle a choisi ses principes de choix (P1).
- (8) Mais alors, si l'on veut dire d'une personne qu'elle structure elle-même ce qu'elle est, psychologiquement parlant, elle doit aussi avoir choisi les principes de choix (P2) sur la base desquels elle a choisi ses principes de choix (P1).
- (9) Et ainsi de suite. Nous sommes maintenant pris dans une régression infinie. Le fait de structurer soi-même ce que nous sommes, psychologiquement parlant, semble impossible car cela exige une série infinie de choix de principes de choix.

---

<sup>37</sup> Strawson, dans son article, attaque le modèle de libre arbitre de Kane qui, comme celui d'Ekstrom, postule que c'est le caractère qui est formé de manière indéterminée. Mon argument a été formulé spécialement pour le modèle d'Ekstrom.

(10) Le modèle d'Ekstrom est donc invalide, car il est impossible de structurer soi-même ce que nous sommes, psychologiquement parlant, et ce, à aucun égard.

L'argument de base de Strawson est un puissant argument contre le modèle d'Ekstrom. La régression présentée ci-haut met en évidence qu'il n'est pas véritablement possible de structurer qui nous sommes. En effet, il est indéniable que ce que nous sommes, initialement, est déterminé par des facteurs génétiques et environnementaux. En utilisant l'expression « ce que nous sommes, initialement », je fais référence à l'ensemble des désirs, des croyances, des valeurs et des idéaux qui nous constituent et qui n'ont pas été soumis à une évaluation de notre part. En raison de ce bagage mental de départ, qui n'a pas été déterminé par nous, il est difficile de voir comment il est possible de structurer soi-même qui nous sommes au sens fort où l'exige Ekstrom. L'idée générale, ici, n'est pas de soutenir qu'il est impossible de changer qui nous sommes, mais plutôt qu'il est impossible de changer qui nous sommes à partir de principes de choix qui ne dépendraient que de nous-mêmes, de manière à échapper au déterminisme. En effet, les stimulations extérieures reçues durant notre développement, que ce soit par l'éducation que nous avons reçue et la culture dans laquelle nous avons grandi, laissent une marque indélébile sur qui nous sommes et il est impossible, contrairement à ce que croit Ekstrom, d'y remédier.

### **3.2.2 La faculté évaluative échappe-t-elle au déterminisme ?**

Comment Ekstrom peut-elle échapper à la régression présentée dans l'argument de base ? Il est important de rappeler qu'Ekstrom ne prétend pas que nous sommes capables de structurer entièrement qui nous sommes. En effet, selon elle, nous pouvons structurer notre caractère, c'est-à-dire choisir nos préférences et nos adhésions, et c'est uniquement cette capacité à choisir soi-même nos préférences et nos adhésions qui importe en ce qui a trait au libre arbitre. La notion de caractère d'Ekstrom est donc beaucoup plus restreinte que la notion ordinaire de caractère, qui inclut, par exemple, certains traits de personnalité, comme le fait d'être colérique, généreux ou orgueilleux. Mais qu'y a-t-il de particulier avec les préférences et les adhésions ? Autrement dit, pourquoi serions-nous capables de choisir nos préférences et nos adhésions, contrairement aux autres éléments qui constituent qui nous sommes ? Selon Ekstrom, les préférences et les adhésions sont des états mentaux bien particuliers qui sont formés par un processus d'évaluation critique. Ce processus d'évaluation, quant à lui, opère en fonction de la faculté évaluative de l'agent, une faculté qu'Ekstrom inclut dans sa notion de soi. La faculté évaluative permet, selon Ekstrom, de revenir sur nos désirs

et sur nos croyances, et ainsi de former des préférences et des adhésions, et ce sans être déterminés à les former. Si Ekstrom peut prétendre échapper au déterminisme et à la régression présentée dans l'argument de base, c'est donc essentiellement grâce à la faculté évaluative. En effet, même si nous n'avons pas choisi, initialement, nos désirs et nos croyances, notre faculté évaluative nous permet, en théorie, de retourner examiner ceux-ci et ultimement de les rejeter ou de les entériner rationnellement, et ainsi d'échapper au déterminisme (en ce qui a trait aux préférences et aux adhésions).

Mais comment opère notre faculté évaluative ? Ekstrom définit la faculté évaluative comme la capacité à former des préférences et des adhésions à la lumière de notre conception du bien et du vrai (Ekstrom 2002, p. 126). L'exercice de notre faculté évaluative est donc relatif à notre conception du bien et du vrai. Cette conception du bien et du vrai, la choisissons-nous ? Encore une fois, il semble que non. La régression présentée dans l'argument de base se représente ici dans toute sa force. Pour que notre faculté évaluative opère, nous devons déjà être pourvus d'une conception du vrai et du bien, sans quoi la faculté évaluative ne peut simplement pas s'exercer, par définition. A priori, il est donc difficile de voir comment notre faculté évaluative pourrait nous permettre d'échapper au déterminisme, comme le prétend Ekstrom. Pour assoir son modèle, Ekstrom doit nous donner de bonnes raisons de croire que notre faculté évaluative a le pouvoir d'échapper au déterminisme. Or, la seule raison qu'elle nous offre est notre expérience interne : « We know from the inside what it is like to deliberate among competing desires or courses of action and to exercise our [evaluative] faculty to decide to form one particular preference rather than another, without being necessitated by prior factor to decide as we do. » (Ekstrom 2000, p. 127) Ces « preuves internes », acquises par introspection, sont-elles fiables ? Il y a de bonnes raisons d'en douter. Nous nous méprenons souvent sur nos impressions internes. Et s'il s'agissait seulement d'une illusion, d'une création de l'esprit ? Qui plus est, l'expérience interne varie fréquemment selon les individus. Dans un débat aussi complexe et conceptuellement chargé que celui sur le libre arbitre, les modèles doivent reposer sur des bases plus solides que certaines données introspectives. En l'absence de preuves convaincantes que l'exercice de notre faculté évaluative introduit une indétermination dans nos délibérations, il vaut mieux rejeter le modèle d'Ekstrom.

Dans cette section sur l'argument de la chance, j'ai évalué la critique selon laquelle les modèles libertariens sont incohérents parce qu'ils introduisent de la chance dans la

chaîne causale des actions. La réponse d'Ekstrom à l'argument de la chance étant convaincante, nous ne pouvons conclure que les modèles incompatibilistes sont incohérents en ayant recours à l'indéterminisme. Faire appel à la chance pour attaquer les modèles libertarianistes est une stratégie largement utilisée tant par les sceptiques que par les compatibilistes. Nous avons constaté, toutefois, que cet appel à la chance ne peut pas se faire avant que la notion de chance n'ait été elle-même précisée. Ma deuxième tentative évitait ce problème. J'ai montré que le libre arbitre libertarianiste, s'il exige que l'on soit capable de structurer nous-mêmes notre caractère, est impossible. Le modèle d'Ekstrom, bien qu'il soit l'un des plus convaincants de la littérature libertarianiste, ne réussit pas à conférer aux agents un véritable contrôle sur leur caractère. Pour cette raison, Ekstrom échoue à résoudre ce qu'elle appelle le « problème du contrôle ». Dans la prochaine section, j'examinerai davantage l'exigence indéterministe des libertarianistes. Plus précisément, je soutiendrai que l'indéterminisme exigé par les libertarianistes n'existe pas.

### 3.3 Le quasi-déterminisme

Aujourd'hui, rares sont ceux qui défendent l'argument du déterminisme sous sa forme classique, c'est-à-dire en défendant la vérité d'un déterminisme universel. Il en est ainsi principalement parce que l'interprétation standard de la mécanique quantique (celle de Copenhague) met grandement en doute la vérité d'un tel déterminisme. Heureusement pour les sceptiques, nul besoin de défendre un déterminisme universel pour s'attaquer au libre arbitre libertarianiste: « the [libertarian] incompatibilist must believe not only that free will entails that there are undetermined events; he must believe that free will entails that there are undetermined events that shape our behaviour » (van Inwagen 1986, p. 126). Les libertarianistes ne peuvent pas se contenter d'un indéterminisme subatomique, ils ont besoin d'un indéterminisme au niveau de l'action. Or, selon Ted Honderich, nous sommes entièrement déterminés au niveau de notre existence. Qu'est-ce que cela veut dire ? L'idée générale est que l'indéterminisme, qui n'existe véritablement qu'au niveau subatomique, n'engendre pas un indéterminisme au niveau neuropsychologique (Honderich 2002). Cette thèse est appelée « quasi-déterminisme » (*near determinism*). Le quasi-déterminisme constitue une menace évidente pour les incompatibilistes. En effet, ces derniers ont besoin, dans leur modèle, d'un moment *pertinent* d'indétermination dans la chaîne causale menant à l'action. Dans cette section, j'attaquerai les modèles libertarianistes de libre arbitre en soutenant qu'il n'existe aucune preuve

concluante de l'existence d'un indéterminisme au niveau de l'action. Commençons par examiner les raisons qu'offre Honderich de croire au quasi-déterminisme.

Honderich considère, d'un côté, que les neurosciences donnent de bonnes raisons de croire qu'il existe des lois neuropsychologiques et qu'elles sont déterministes (Honderich 2002, p. 68). Il soulève, en exemple, une variété d'expériences. L'une d'entre elles concerne la mémoire et la stimulation de certaines régions du cerveau. La stimulation d'une partie très précise du cortex à l'aide d'électrodes provoque, chez les patients de l'expérience, un *flashback* particulier, et chaque fois que la même stimulation est faite chez le même patient, le même *flashback* est provoqué. Il semble, dans ce cas, que des événements neuronaux (stimulation du cortex) engendrent toujours les mêmes événements mentaux (*flashback*). Le même type d'expérience peut être effectué à propos des émotions, de la douleur et du plaisir. Les facultés supérieures ne semblent pas échapper au déterminisme. En effet, il semble que l'endommagement de certaines parties de l'aire de Broca (une aire du cerveau reconnue pour ses liens étroits avec le langage) entraîne systématiquement une perte de certaines facultés langagières<sup>38</sup>. Toutes ces expériences, bien qu'elles soient loin d'être décisives, nous donnent des raisons de croire qu'il y a des lois neuropsychologiques déterministes<sup>39</sup>.

D'un autre côté, la physique quantique, contrairement aux apparences, ne nous donne aucune raison de croire que les lois neuropsychologiques sont indéterministes. La physique quantique, selon son interprétation standard, soutient que les lois de la nature sont indéterministes au niveau subatomique<sup>40</sup>. Cet indéterminisme au niveau subatomique, selon les incompatibilistes, peut engendrer un indéterminisme au niveau neuropsychologique. Les incompatibilistes fondent leur théorie sur cette possibilité. Malgré la *possibilité* d'un tel trans-

---

<sup>38</sup> Honderich ne donne aucune référence lorsqu'il fait appel à ces expériences. Il ne m'est donc pas possible de vérifier si ces expériences sont fiables. Espérons qu'elles le soient.

<sup>39</sup> Honderich défend trois types de déterminisme à la fois : le *mind-body determinism*, l'*initiation determinism* et l'*action determinism*. Les expériences mentionnées ci-haut supportent le *mind-body determinism*, que j'ai préféré appeler « déterminisme neuropsychologique ». Voir Honderich (2002, p. 63).

<sup>40</sup> Honderich doute fortement de l'interprétation standard de la physique quantique. Selon lui, l'interprétation de Copenhague, acceptée depuis plus de 75 ans, n'est supportée par aucune preuve expérimentale (Honderich 2002, p. 73). L'examen des différentes interprétations de la physique quantique n'est pas à ma portée. Je me contenterai d'expliquer la critique d'Honderich à propos du transfert d'indéterminisme au niveau neuropsychologique.

fert d'indéterminisme, Honderich ne voit simplement aucune raison de croire que ce soit effectivement le cas : « What about the evidence for that? Why have we not noticed one of these chance events? Why has a spoon not levitated before now, when the random lurches of little events within it all happened to combine in the right way? » (Honderich 2002, p. 75). La réponse commune à cette objection est de soutenir que de tels événements sont tellement improbables qu'il ne faut pas s'attendre à pouvoir les observer. Honderich trouve cette réponse insatisfaisante. Avec raison, cette objection n'aidera certainement pas les libertariens, qui ne souhaitent pas accorder aux actions libres une existence rarissime. De plus, même en admettant que l'interprétation standard de la physique quantique est correcte, il existe une différence d'échelle impressionnante entre le niveau subatomique ( $10^{-15}$  m) et le niveau neuronal ( $10^{-5}$  m). En l'absence de données prouvant le contraire, tout porte à croire que l'indéterminisme au niveau subatomique se dissipe complètement au niveau de notre existence. C'est à partir de ces quelques considérations qu'Honderich écrit: « At the ordinary level of choices and actions, and even ordinary electrochemical activity in our brains, [deterministic<sup>41</sup>] causal laws govern what happens » (Honderich 2002, p. 5).

### 3.3.1 L'appui des sciences physiques au libertarianisme

Pour augmenter la plausibilité d'un indéterminisme au niveau neuropsychologique, il est possible de joindre la physique quantique aux théories du chaos : « [...] chaos may be the magic that provides for human freedom. » (Garson 1995, p. 366) Si la physique quantique à elle seule ne permet pas de réaliser le modèle libertarien de libre arbitre, peut-être que la conjonction de celle-ci avec les théories du chaos aura plus de succès. C'est l'approche que favorise Ekstrom. En effet, quand on lui demande comment son modèle indéterministe de libre arbitre est réalisé, Ekstrom répond : « Several recent theorists have done interesting exploratory work on the matter of how indeterminism at the quantum level might be magnified if the brain is a chaotic system<sup>42</sup>. » (Ekstrom 2000, p. 124) On pourrait croire, à partir de l'énoncé d'Ekstrom, que c'est parce que le cerveau est un système chaotique et que les systèmes chaotiques sont indéterministes qu'il y a un indéterminisme au niveau neuropsycholo-

---

<sup>41</sup> Lorsque Honderich parle de lois causales, il parle de lois causales déterministes. En doutant de l'existence de l'indéterminisme (même au niveau quantique), Honderich se permet, à mon avis, de ne pas le préciser.

<sup>42</sup> Ekstrom fait référence ici à Kane (1996, pp. 128-130) et à Jesse Hobbs (1991).

gique. En réalité, la position d'Ekstrom ne doit pas être interprétée de cette manière, car il est maintenant admis que les théories du chaos sont entièrement déterministes (McFee 2001; Hobbs 1991; Kane 1996; Bishop 2011). Plus précisément, on dira que les théories du chaos sont indéterministes au sens épistémologique, c'est-à-dire qu'il n'est pas possible de *connaître* toutes les variables qui entrent en jeu dans un système chaotique. Il en résulte une impossibilité à *prédire* comment un système chaotique va réagir. Toutefois, les théories du chaos sont ontologiquement déterministes, en ce sens que les événements, dans un système chaotique, sont soumis à des lois déterministes (qu'on ne peut pas connaître vu la complexité des systèmes chaotiques). Cela est admis, autant par Honderich que par les auteurs auxquels Ekstrom fait référence. Il ne s'agit donc pas d'une solution satisfaisante. Le fond de l'argument ressemble plutôt à ceci: « Since chaotic events are exquisitely dependent on the very small, the doings of sub-atomic particles in a very small region of space-time may be reflected in massive differences in global system 'behavior. » (Garson 1995, p. 368) Il semble adéquat de s'imaginer la thèse ainsi : les systèmes chaotiques permettent à l'indéterminisme subatomique d'être amplifié jusqu'au niveau de notre existence, à l'instar d'une réaction en chaîne lors d'une explosion atomique. Comme nous le verrons sous peu, les systèmes chaotiques n'ont besoin que d'une infime stimulation pour générer d'énormes répercussions sur le système en entier. En faisant appel aux théories du chaos, Ekstrom n'est plus obligée de défendre l'existence d'un indéterminisme généralisé dans le cerveau. Le modèle d'Ekstrom pourra, prétendument, se contenter d'un nombre très restreint d'événements indéterministes. Le chaos dans le cerveau se chargera d'amplifier le tout jusqu'à ce que l'indéterminisme ait un impact sur le comportement humain. Je crois pouvoir, de manière charitable, prêter le raisonnement suivant à Ekstrom :

- (1) Si le cerveau est un système chaotique, alors l'indéterminisme au niveau subatomique peut s'élever au niveau neuronal.
- (2) Le cerveau est un système chaotique.
- (3) L'indéterminisme au niveau subatomique peut s'élever au niveau neuronal.

Cette inférence, évidemment valide, repose sur la plausibilité des prémisses (1) et (2). Dans les paragraphes suivants, j'examinerai chacune des deux prémisses. La plausibilité des deux prémisses repose essentiellement sur des considérations empiriques. En examinant la plausibilité de chacune de ces prémisses, j'espère mettre au clair le manque de preuves empiriques entourant la réalisation neuronale du libre arbitre libertarien d'Ekstrom.

### 3.3.2 Un manque de preuves empiriques

Le cerveau est-il un système chaotique ? Pour répondre à cette question, il faut d'abord savoir ce qu'est un système chaotique. Aux fins de l'argument, nous dirons qu'un système est chaotique si et seulement si un changement minime dans les conditions initiales du système peut se développer de manière exponentielle et engendrer d'énormes changements au final (Kane 1996, p. 129). L'idée peut être comprise, plus simplement, en faisant référence à l'effet papillon. Dans un système climatique très chaotique, par exemple, le battement d'ailes d'un papillon pourrait causer la formation d'une tornade à l'autre bout de la planète (Bishop 2009). Selon Hobbs, presque tous les systèmes dissipatifs et apériodiques sont chaotiques (Hobbs 1991, p. 143). Ici, « dissipatif » est synonyme de « système ouvert », c'est-à-dire un système où il y a un échange d'énergie ou de matière avec d'autres systèmes. Bishop (2009) définit l'apériodicité d'un système ainsi: « Aperiodic behavior means that the system variables never repeat values in any regular fashion ». « Apériodique » est synonyme de « irrégulier », en ce sens que ce type de système ne réagit pas selon un modèle (*pattern*) bien défini. Ainsi, considérant que tous les systèmes biologiques sont des systèmes dissipatifs apériodiques, il semble que le système nerveux soit lui aussi un système chaotique. Si tel est le cas, la sécrétion d'une seule molécule, ou l'envoi d'un seul électron à une synapse, pourrait avoir des répercussions significatives sur le système nerveux en entier, et donc sur le comportement humain. Toutefois, la question empirique de savoir si le cerveau est un système chaotique est, encore aujourd'hui, sujette à de grands débats (Bishop 2009) et il n'est pas encore possible de tirer de conclusion satisfaisante à ce sujet (Bishop 2011, p. 92). L'avenir nous dira, semble-t-il, si la prémisse (2) de l'argument prêté à Ekstrom est vraie, c'est-à-dire si le cerveau est un système chaotique, ou du moins, aussi chaotique que semble le croire Ekstrom.

Les systèmes chaotiques permettent-ils d'amplifier l'indéterminisme au niveau subatomique ? Admettons que le système nerveux soit un système chaotique. Si tel est le cas, il faut tout de même, minimalement, que l'indéterminisme au niveau subatomique se réalise sur au moins un atome ou une molécule du système nerveux. Ce faisant, l'indéterminisme au niveau subatomique pourrait avoir des répercussions au niveau neuropsychologique. Bishop le décrit bien:

Along these lines, suppose the patterns of neural firings in the brain correspond to decision states. Chaos could amplify quantum events, causing a single neuron

to fire that would not have fired otherwise. If the brain (a macroscopic object) is also in a chaotic dynamical state, which makes it sensitive to small perturbations, this additional neural firing, small as it is, would then be further amplified to the point where the brain states would evolve differently than if the neuron had not fired. In turn these altered neural firings and brain states would carry forward such quantum effects that affect the outcomes of human choices. (Bishop 2011, p. 92)

Pour qu'il y ait amplification, il doit y avoir quelque chose à amplifier. Est-ce le cas ? Certains auteurs soutiennent que oui lorsqu'il est question de friction, mais uniquement quand les effets quantiques lors de la friction sont assez grands pour briser certains liens moléculaires et que l'amplification se fait assez rapidement (Hobbs 1991; Bishop 2011). Malheureusement, on ne connaît pas les contraintes d'amplification dans le cas du cerveau (Bishop 2011, p. 93). Nous avons, encore ici, affaire à des enjeux proprement empiriques. Malheureusement, les théories du chaos sont encore jeunes et il n'existe pas de résultats concluants à propos des conditions d'amplification de l'indéterminisme à travers le système possiblement chaotique du cerveau. Il n'est pas encore possible de savoir si la prémisse (1) de l'argument prêté à Ekstrom est vraie.

Le libertarianisme, en exigeant la présence d'un moment d'indétermination dans la chaîne causale menant aux actions libres, n'a d'autres choix que de nier le quasi-déterminisme. C'est le prix à payer pour satisfaire leur intuition incompatibiliste. Dans cette section, j'ai fait valoir que le libertarianisme ne possédait aucune preuve concluante de l'existence d'un indéterminisme au niveau de l'action. Nous avons constaté que faire appel à l'indéterminisme quantique est loin d'être évident. D'un côté, en admettant que l'interprétation standard de la physique quantique est correcte, nous sommes confrontés à certaines preuves d'un déterminisme au niveau neuropsychologique. D'un autre côté, nous n'avons aucune raison de croire que l'indéterminisme quantique peut avoir un impact sur l'action. De plus, nous ne savons presque rien sur le possible système chaotique que serait le cerveau. Ces considérations ne sont pas suffisantes pour démontrer que les modèles incompatibilistes sont voués à l'échec. Toutefois, le caractère très programmatique des évidences empiriques sur lesquelles ils reposent empêche de conclure à leur validité dans l'état actuel des connaissances. Jumelées à l'argument précédent selon lequel l'indéterminisme ne permet pas d'assurer le contrôle que doit avoir un agent pour être libre, les difficultés formulées dans cette section donnent de bonnes raisons de croire que les modèles libertarianistes doivent être abandonnés au profit d'une solution moins hasardeuse.

### 3.4 Conclusion

Les objectifs de ce chapitre étaient de proposer une critique convaincante du libertarianisme. Premièrement, j'ai essayé de soutenir que les libertariens, en exigeant de l'indéterminisme dans la chaîne causale menant aux actions libres, introduisaient de la chance dans cette même chaîne causale, de telle sorte qu'il est impossible qu'un agent soit en contrôle de ses actions (van Inwagen 2000). Après avoir examiné la réponse d'Ekstrom (2011) à cette objection, j'ai jugé que cette critique ne tenait pas la route (du moins, contre le modèle d'Ekstrom). Je crois avoir démontré, toutefois, que le libre arbitre soutenu par Ekstrom, en exigeant que l'on soit capable de structurer nous-mêmes notre caractère, est impossible. Deuxièmement, j'ai défendu le quasi-déterminisme, la thèse selon laquelle nous sommes déterminés au niveau de notre existence. La vérité du quasi-déterminisme mine grandement la plausibilité des modèles libertariens. Par la suite, j'ai mis en évidence le manque de preuves empiriques concernant la réalisation neuropsychologique des modèles libertariens qui font appel aux théories du chaos pour augmenter la plausibilité de leur théorie. Parce que l'indéterminisme ne contribue pas au contrôle que peut avoir un agent sur ses actions et parce que le quasi-déterminisme est vrai, j'ai jugé qu'il vaut mieux rejeter les modèles libertariens du libre arbitre.

Après avoir attaqué le compatibilisme et le libertarianisme, nous sommes fin prêts à explorer les voies du scepticisme. Dans le prochain chapitre, je prendrai ainsi la défense d'un modèle sceptique de libre arbitre, l'illusionnisme de Smilansky (2000). D'après Smilansky, le libre arbitre n'existe pas véritablement, il ne s'agit que d'une illusion. Cette illusion, selon Smilansky, est nécessaire au maintien des sociétés, et c'est pour cette raison qu'il vaut mieux continuer à croire (à tort) que nous sommes libres.

## Chapitre 4 : Pour le scepticisme

### 4.0 Introduction

Dans les deux chapitres précédents, nous avons vu que le compatibilisme et le libertarianisme font face à des difficultés de taille. Les modèles compatibilistes ont été rejetés parce que le libre arbitre n'est pas compatible avec le déterminisme et parce que les modèles compatibilistes ne réussissent pas à rendre compte de façon satisfaisante des cas de manipulation. Les modèles libertarianistes, quant à eux, ont été rejetés, d'une part parce qu'il semble impossible de structurer soi-même son caractère (une idée défendue par plusieurs libertarianistes contemporains) et, d'autre part, parce que nous n'avons aucune bonne raison de croire en l'existence d'un indéterminisme au niveau de l'action. Quelle place reste-t-il pour le libre arbitre, si celui-ci est menacé à la fois par le déterminisme et par l'indéterminisme ? Dans de telles circonstances, la voie est grande ouverte pour le scepticisme. Mais si le libre arbitre n'existe pas, qu'en est-il de la responsabilité morale ? Qu'en est-il de la justice ? Sommes-nous justifiés de punir un meurtrier si celui-ci n'était pas libre d'assassiner sa victime ? Le scepticisme à propos du libre arbitre fait face à des difficultés bien particulières, à des difficultés d'ordre normatif. Comme pour le compatibilisme et le libertarianisme, il existe plusieurs façons de développer un modèle sceptique à propos du libre arbitre. Afin d'exemplifier le scepticisme, et afin d'augmenter la plausibilité d'une telle thèse, je présenterai ce que je crois être l'un des meilleurs modèles sceptiques contemporains, l'illusionnisme de Saul Smilansky (2000). Dans *Free Will and Illusion*, Smilansky prétend éviter la catastrophe d'un monde dans lequel personne ne serait responsable de ses actions en développant un modèle sceptique qui fait appel à l'illusion du libre arbitre. Ce dernier chapitre sera consacré à l'examen de ce modèle. Le modèle de Smilansky est extrêmement complexe et pourrait, à lui seul, faire l'objet d'un mémoire entier. J'essaierai d'être fidèle à sa pensée sans entrer dans tous ses détails. Je tenterai de montrer comment l'illusionnisme surmonte les difficultés normatives du scepticisme à propos du libre arbitre, tout en évitant les difficultés spécifiques au compatibilisme et au libertarianisme. Ce faisant, j'espère convaincre mon lecteur que le scepticisme à propos du libre arbitre est une avenue qui mérite d'être considérée sérieusement.

## 4.1 Un modèle normatif

Avant d'examiner la théorie illusionniste de Smilansky, il m'est nécessaire de faire une remarque préliminaire. Pour comprendre la position de Smilansky, il faut savoir que le libre arbitre occupe, selon lui, une place centrale en ce qui a trait à la responsabilité morale et à la justice. En effet, le modèle de Smilansky s'insère dans un courant traditionnel qui remonte à Aristote qui, dans *l'Éthique à Nicomaque*, soutient que le libre arbitre est une condition nécessaire à la responsabilité morale. Smilansky lie ainsi logiquement la notion de libre arbitre à celle de responsabilité morale. Comme je l'ai mentionné au chapitre 1, j'ai préféré mener les réflexions de ce mémoire en évitant de lier la responsabilité morale et le libre arbitre. Smilansky, quant à lui, est convaincu qu'il existe un lien très fort entre les deux notions<sup>43</sup>. Il me faut donc m'y arrêter un peu maintenant pour présenter adéquatement son approche.

L'importance de la responsabilité morale et de la justice pour Smilansky se manifeste dans la thèse qu'il appelle la « conception centrale du libre arbitre » (*Core Conception of Free Will*). Selon cette conception du libre arbitre, la fonction d'un modèle de libre arbitre est de fonder l'éthique et la justice, au sens où le libre arbitre sert à dissiper l'aspect arbitraire de ces deux domaines :

The broad intuition behind the Core Conception is that it is arbitrary—and thus prima facie morally problematic—if people are worse or better off on account of factors not within their control or 'up to' them. Control allows the shrinking of the morally arbitrary, impinging the moral on the arbitrary. To begin with, people cannot truly deserve to be worse or better off except as a consequence of their own actions, freely performed: how can they deserve if they had not freely done something to be deserving? This intuition is strongest in forms of evaluation such as blame, which inherently assume that the person being blamed deserves to be because she was responsible, had the blameworthy factor to some extent under her control. For if a person is blamed for something that was not in fact up to her she does not deserve blame, and this would seem to be a paradigm of injustice. (Smilansky 2000, p. 16)

La conception centrale du libre arbitre est fondée sur l'intuition qu'il est injuste de punir ou de blâmer un individu s'il n'est pas en contrôle de ses actions. La conception centrale du libre arbitre constitue le principal critère de Smilansky afin de déterminer si un modèle de libre

---

<sup>43</sup> Il dira même que la responsabilité et le libre arbitre semblent liés de manière « presque analytique » (Smilansky 2000, p. 16).

arbitre est acceptable. En d'autres mots, on dira qu'un modèle de libre arbitre doit réussir à dissiper l'aspect arbitraire de l'éthique et de la justice, et s'il y échoue, le modèle doit être rejeté. Le libre arbitre, pour Smilansky, a donc une portée essentiellement normative. Nous aurons l'occasion d'examiner plus en détail la conception centrale du libre arbitre à travers l'argumentation de Smilansky.

#### 4.1.1 La communauté de responsabilité

Maintenant que la conception centrale du libre arbitre a été brièvement exposée, examinons comment cette conception se manifeste dans nos sociétés. Smilansky appelle les sociétés qui sont fondées sur la conception centrale du libre arbitre des « communautés de responsabilité » (Smilansky 2000, p. 83). Dans une communauté de responsabilité, les interactions sociales et les pratiques institutionnelles présupposent que les agents sont responsables de leurs actions<sup>44</sup>. Par exemple, si nous blâmons un ami qui n'a pas tenu sa promesse de nous rembourser de l'argent, c'est parce que nous jugeons qu'il est responsable de ses actions et qu'il est en contrôle des décisions et des actions qui l'ont amené à ne pas respecter sa promesse. Si nous donnons des prix aux grands scientifiques de nos sociétés, c'est parce que nous jugeons qu'ils méritent d'être récompensés pour leurs actions, et donc qu'ils sont responsables de celles-ci. Si nous trouvons qu'il est injuste de punir un enfant et de l'envoyer en prison, c'est parce que nous jugeons qu'il n'est pas pleinement responsable de ses actions. Bref, tout porte à croire que nous vivons présentement dans une communauté de responsabilité. Mais alors, comment réagiraient les membres de notre société s'ils apprenaient que le libre arbitre n'existe pas et que personne, ultimement, n'est responsable de ses actions ? Pour Smilansky, vivre dans une communauté de responsabilité est un impératif moral (Smilansky, p. 160). La réalisation complète de l'inexistence du libre arbitre minerait le fondement même des interactions sociales et des pratiques institutionnelles, ce qui serait évidemment dramatique. Le libre arbitre, ou plutôt l'illusion du libre arbitre puisque Smilansky maintient que le libre arbitre n'existe pas, joue ainsi un rôle central dans le maintien de nos sociétés, et il est essentiel, selon Smilansky, d'en rendre compte : « Within our lives, insofar

---

<sup>44</sup> La notion de communauté de responsabilité mise en place par Smilansky est étroitement liée au compatibilisme. Parce qu'il s'agit d'un lien complexe et peu pertinent dans le cadre de mon argumentation, j'ai préféré ne pas l'explicitier.

as they are affected by the issue of free will, illusion is descriptively central and normatively necessary. » (Smilansky 2000, p. 1).

## 4.2 La critique de Smilansky

Il convient, maintenant, d'examiner l'analyse que fait Smilansky du compatibilisme, du libertarianisme et du déterminisme dur (la thèse selon laquelle la responsabilité morale n'existe pas). Cet examen est nécessaire, dans la mesure où l'illusionnisme de Smilansky est motivé, en grande partie, par l'échec de ces trois approches. L'objectif, ici, est de donner un aperçu de la critique de Smilansky, en tissant des liens avec les critiques que j'ai adressées dans ce mémoire aux défenseurs du libre arbitre. La critique générale de Smilansky se présente ainsi : aucune des trois approches standards – le compatibilisme, le libertarianisme et le déterminisme dur – ne réussit à conférer le contrôle qui est nécessaire pour dissiper l'aspect arbitraire de l'éthique et de la justice ainsi que l'exigent la conception centrale du libre arbitre et l'intuition essentielle sur laquelle elle se fonde.

Pourquoi le compatibilisme échoue-t-il à cette tâche ? Smilansky résume la situation ainsi : les compatibilistes ne réussissent pas à conférer un contrôle *ultime* aux agents (Smilansky 2000, p. 47). D'ailleurs, ils ne prétendent pas le faire<sup>45</sup>. La notion de contrôle ultime exige qu'un agent soit à la source de la chaîne causale menant à ses actions. C'est une notion que privilégient généralement les libertarianistes et les déterministes durs. Selon Smilansky, ce contrôle ultime est nécessaire pour satisfaire la conception centrale du libre arbitre : « it is unjust to decide, in a "arbitrary way", to stop going "as far back" as we can, when looking for people's opportunities and abilities » (Smilansky 2000, p. 82). Qu'est-ce que cela veut dire ? Nous avons vu, au chapitre 2, que certains modèles compatibilistes étaient historiques, au sens où ils s'intéressaient à l'histoire causale des actions. Par contre, les modèles compatibilistes ne peuvent pas s'intéresser à toute l'histoire causale des actions, sans quoi ils seraient confrontés, ultimement, à une cause qui déterminerait l'action sans que l'agent ait de

---

<sup>45</sup> Il est important, je crois, de rappeler que le compatibilisme n'a jamais eu la prétention de conférer un contrôle ultime aux agents. Smilansky l'exprime bien lorsqu'il écrit que les compatibilistes sont des « non-demanding optimist[s] » (Smilansky 2000, p. 3), en ce sens qu'ils ne sont pas très exigeants quant aux conditions d'existence du libre arbitre. Dans le même ordre d'idées, les libertarianistes sont qualifiés de « demanding optimist[s] » et les déterministes durs sont qualifiés de « demanding pessimist[s] » (Smilansky 2000, p. 4).

contrôle sur cette cause. En effet, si le déterminisme est vrai, la séquence causale menant à une action ne commence pas lors d'une délibération (comme pouvait le prétendre Ekstrom, par exemple, en localisant un indéterminisme durant les délibérations à propos de nos préférences). La chaîne causale qui précède une action, dans un monde déterministe, n'a pas de début défini. Si le compatibilisme est « arbitraire », c'est parce qu'il pose une limite à l'investigation des chaînes causales menant à nos actions, et que cette limite ne semble pas justifiable dans un monde déterministe. Frankfurt, par exemple, fondait le libre arbitre sur les désirs du second ordre, mais ne s'intéressait pas aux causes de ces désirs du second ordre. Fischer et Ravizza, de leur côté, arrêtaient leur investigation aux mécanismes ayant causé l'action, ou plus précisément, à certaines croyances à propos de ces mécanismes. Or, dans un monde déterministe, limiter son investigation comme le font les compatibilistes est inadmissible.

Qu'en est-il, d'autre part, du libertarianisme ? Smilansky soutient qu'il s'agit d'un projet impossible et incohérent (Smilansky 2000, p. 48). Qu'est-ce qui l'amène à cette conclusion ? Selon Smilansky, afin de satisfaire la conception centrale du libre arbitre, les libertarianistes doivent respecter un principe qu'il appelle le « principe de l'attribution unique » (*the Principle of Sole Attribution*). Ce principe est formulé ainsi: « Any feature F due to which a person deserves something S in the libertarian free will-dependent sense must, in the normatively relevant respects, be solely attributable to the person or to the pertinent aspect A of the person. » (Smilansky 2000, p. 62). « F » fait référence à une décision, à un acte de volonté ou à une action, « S » fait référence à un blâme, à des louanges, ou à une punition quelconque et « A » fait référence à certains traits de caractère ou à certaines capacités morales, par exemple. L'idée générale est qu'un agent mérite d'être blâmé ou louangé pour une action seulement si l'action peut être uniquement attribuée à l'agent. On dira qu'une action peut être uniquement attribuée à un agent si et seulement si cette action n'a pas été causée par quelque chose qui est hors du contrôle de cet agent (p. 63). Par exemple, il serait absurde pour un libertarianiste de blâmer un individu violent si les actions violentes de ce dernier ont été causées, ultimement, par son bagage génétique et par certains événements traumatisants de son enfance.

Smilansky soutient que les libertarianistes cherchent tous à respecter le principe de l'attribution unique, et que c'est à partir de ce principe qu'ils critiquent le compatibilisme. Smilansky le démontre en examinant les modèles libertarianistes de C.A. Campbell (1967), de

Roderick Chisholm (1982), de David Wiggins (1973), de Richard Sorabji (1980) et de Robert Kane (1996). L'analyse qu'il fait de ces modèles est longue et nécessiterait trop d'espace pour être présentée dans ce mémoire. On se rappelle toutefois qu'Ekstrom cherchait elle aussi à respecter ce principe : « free actions, on [my] view, are not random or arbitrary; they are directly under the agent's control in being determined by and only by the agent » (Ekstrom 2002, p. 127). Selon Ekstrom, les actions libres peuvent être attribuées uniquement à l'agent, car c'est l'agent qui détermine de façon nécessaire ses actions, l'indéterminisme étant localisé uniquement durant la formation de notre caractère. La critique de Smilansky consistera à dire que les libertariens ne respectent pas le principe de l'attribution unique, parce que la formation de notre caractère ne peut pas être attribuée uniquement à nous. Smilansky fait d'ailleurs lui aussi appel à l'argument de base de Strawson pour démontrer ce point. Parce que les conditions posées sur le libre arbitre par le libertarisme ne peuvent être remplies, Smilansky en conclut que la notion de libre arbitre que propose cette approche est impossible et incohérente.

Finalement, Smilansky juge que le déterminisme dur, la thèse selon laquelle la responsabilité morale n'existe pas, n'est pas défendable non plus (Smilansky 2000, p. 74). Ce jugement peut être expliqué en juxtaposant le déterminisme dur et la conception centrale du libre arbitre. D'un côté, la conception centrale du libre arbitre nous dit que la fonction d'un modèle de libre arbitre est de fonder l'attribution de responsabilité morale. D'un autre côté, les déterministes durs admettent qu'il faut un contrôle ultime sur nos actions pour être libre et moralement responsable de celles-ci. Toutefois, la vérité du déterminisme rend impossible l'existence d'un tel contrôle, ce qui rend invalide toute attribution de responsabilité morale. En conséquence, dans la perspective du déterminisme dur, il n'existe aucune différence, en termes de responsabilité morale, entre un schizophrène qui assassine un individu pendant un épisode psychotique et un carriériste qui assassine un de ses collègues pour obtenir une promotion. Dans les deux cas, l'homicide est le résultat nécessaire des événements du passé en conjonction avec les lois naturelles, et ni le schizophrène ni le carriériste ne sont moralement responsables de leurs actions. Devant un tribunal, les deux situations devraient être traitées de la même façon. Or, ce résultat semble injuste pour le schizophrène. Le déterminisme dur, même s'il reconnaît avec raison que personne n'est ultimement en contrôle de ses actions, est donc jugé indéfendable d'un point de vue normatif parce qu'il conduit à la négation de l'existence de la responsabilité morale dans des cas où, intuitivement, nous souhaitons la maintenir.

La position de Smilansky n'est pourtant pas très éloignée de celle du déterminisme dur. Smilansky est d'accord avec les déterministes durs à propos du fait qu'un contrôle ultime est requis pour légitimer l'attribution de responsabilité morale. Il est aussi d'accord avec eux à propos du fait que le libertarianisme ne réussit pas à conférer ce contrôle ultime. Finalement, il est d'accord avec eux à propos du fait que le contrôle conféré par les modèles compatibilistes est trop faible. Le désaccord se joue sur un plan normatif. Les déterministes durs ne voient pas nécessairement de problème à ce que la responsabilité morale n'existe pas. Or, pour Smilansky, l'absence de responsabilité morale est tout simplement dramatique, et il se refuse à accepter un modèle « simpliste » et « moralement insensible » de déterminisme dur (Smilansky 2000, p. 92).

#### 4.2.1 L'hypothèse du monisme

Pourquoi les modèles standards ont-ils fait fausse route ? Selon Smilansky, les trois modèles standards sont tombés dans l'erreur parce qu'ils entérinent tous l'hypothèse du monisme (Smilansky 2000, p. 36). L'hypothèse du monisme nous dit que les attributions de responsabilité morale sont fondées sur un seul type de libre arbitre : ou bien le libre arbitre compatibiliste, ou bien le libre arbitre ultime des libertarianistes. D'un côté, les compatibilistes acceptent l'hypothèse du monisme et soutiennent que le libre arbitre compatibiliste, bien qu'il ne remonte pas les chaînes causales jusqu'à la source ultime des actions, suffit à fonder les attributions de responsabilité morale. D'un autre côté, les libertarianistes et les déterministes durs acceptent l'hypothèse du monisme et soutiennent que c'est uniquement le libre arbitre ultime des libertarianistes qui permet de fonder la responsabilité morale<sup>46</sup>. Qui a raison, en ce qui concerne le lien entre la responsabilité morale et le libre arbitre ? Les compatibilistes, ou les libertarianistes et les déterministes durs ? Un peu tout le monde selon Smilansky. En effet, à son avis, les deux manières d'envisager ce lien mettent en lumière des points importants et il existe une alternative intermédiaire, qui allie des éléments de chacune de ces positions. Comment est-ce possible ? Après tout, les deux conceptions du lien entre responsabilité et libre arbitre sont logiquement inconsistantes entre elles. La solution de Smilansky se résume ainsi : la notion de responsabilité morale est complexe, et pour en rendre pleinement

---

<sup>46</sup> Évidemment, les libertarianistes et les déterministes durs se distinguent ensuite, du fait que les libertarianistes soutiennent que le libre arbitre libertarianiste existe, alors que les déterministes durs soutiennent que le libre arbitre libertarianiste n'existe pas et donc que la responsabilité morale n'existe pas.

compte, nous avons besoin à la fois de la perspective du compatibilisme et de la perspective ultime que partagent le libertarisme et le déterminisme dur. Comment Smilansky articule-t-il cette idée ? Selon Smilansky, les modèles standards ont tous présupposé qu'il existe une seule notion exhaustive de responsabilité morale, et que cette notion pouvait être fondée sur une seule notion exhaustive de libre arbitre. Mais qu'en est-il vraiment ? Ce canevas conceptuel semble plutôt simpliste, et Smilansky nous invite à le rejeter. Le rejet de l'hypothèse du monisme implique qu'il existe plus d'une façon d'envisager le lien entre responsabilité morale et libre arbitre. Il est possible, par exemple, que certaines attributions spécifiques de responsabilité morale nécessitent une notion compatibiliste de libre arbitre, alors que d'autres nécessitent une notion libertariste de libre arbitre : « it may turn out that there is no single or exhaustive notion of moral responsibility » (Smilansky 2000, p. 37). La manière dont s'articulent ces deux façons de concevoir le lien entre responsabilité et libre arbitre est ouverte à interprétation, mais l'idée de fond n'est certainement pas sans intérêt. L'hypothèse du monisme avait certainement l'avantage de simplifier le débat, mais à quel prix ? Smilansky (p. 39) écrit : « The intuitive attraction of the Assumption of Monism is great, but once we cross the "intuitive Rubicon" and move beyond it we see that its parsimony is nothing but false economy. » Lorsqu'il est question du libre arbitre et de la responsabilité morale, Smilansky ne croit pas que nous pouvons nous permettre d'être aveugles à l'une des deux perspectives. Il est peut-être temps d'admettre que ces deux perspectives mettent l'accent sur les deux côtés d'une même médaille.

#### **4.2.2 Un dualisme fondamental**

Le rejet de l'hypothèse du monisme nous place toutefois dans une position inconfortable. En effet, nous devons maintenant embrasser deux perspectives à propos du libre arbitre et de la responsabilité morale, ce qui engendre des problèmes lorsque vient le temps de faire des attributions de responsabilité morale en pratique. Afin d'exemplifier la tension qui existe entre la perspective des compatibilistes et la perspective de ceux qui privilégient la conception libertariste du libre arbitre, examinons le cas des triplets, développé par Smilansky (2000, p. 95) : imaginons trois frères, des triplets, que l'on nommera A, B et C. Le premier frère, A, vient de commettre un vol, alors que le deuxième, B, a résisté à la tentation. Le troisième frère, C, a aussi commis un vol, mais C est un kleptomane diagnostiqué. Comment juger de la responsabilité morale des trois frères ? Si nous prenons la perspective du compatibilisme, il est correct de distinguer A et B d'un côté et C de l'autre. En effet, dans cette

perspective, nous voulons blâmer A pour avoir commis un vol de son plein gré et nous voulons louer B pour avoir résisté à la tentation. C, quant à lui, sera excusé par les compatibilistes, selon des motifs qui différeront d'un modèle à l'autre<sup>47</sup>. Le compatibilisme nous offre donc de bons arguments pour justifier une distinction que nous souhaitons intuitivement faire entre A, B et C. Si nous utilisons une conception libertarienne du libre arbitre, par contre, il est injuste de distinguer A, B et C. En effet, selon la conception libertarienne du libre arbitre, une action sera dite libre seulement si elle peut être attribuée uniquement à un agent. Or, il est impossible qu'une action soit attribuée uniquement à un agent, comme nous l'avons vu plus tôt. Nous devons donc accepter qu'il n'existe aucune différence, en termes de responsabilité morale, entre A, B et C, ainsi que l'affirme la perspective du déterminisme dur. Ce jugement est aussi très intuitif. Après tout, nous agissons tous en fonction de notre caractère, un caractère que nous n'avons pas choisi. Pourquoi, de ce fait, devrions-nous blâmer ou louer l'un ou l'autre des trois frères ?

Nous sommes donc maintenant confrontés à ce que Smilansky appelle un « dualisme fondamental ». Nullement apparenté au dualisme corps-esprit, le dualisme fondamental de Smilansky fait référence à la nécessité théorique d'accepter la validité partielle du compatibilisme et la validité partielle du déterminisme dur. Selon Smilansky, la tension engendrée par ce dualisme fondamental est insoutenable par les communautés de responsabilité :

We cannot live adequately with the two valid sides of the Fundamental Dualism, not with a complete awareness of the absence of libertarian free will. We have to face the fact that there are basic beliefs that morally ought not to be abandoned (or morally ought to be strengthened), although they might destroy each other, or are partially based on incoherent conceptions. (Smilansky 2000, p. 6)

En effet, d'un côté, la vision du compatibilisme nous conduit à porter des blâmes et à adresser des louanges, alors que d'un autre côté, la vision du déterminisme dur nous pousse à ne pas le faire. Selon Smilansky, une bonne partie de nos croyances morales sont fondées sur l'existence du libre arbitre libertarienne. Ces croyances, même si elles sont fondées sur une conception incohérente de libre arbitre, doivent être maintenues à tout prix. En l'absence du

---

<sup>47</sup> Frankfurt, par exemple, nous dirait que C n'est pas moralement responsable du vol, car C n'avait pas le désir de désirer commettre un vol. Fischer et Ravizza, quant à eux, diraient que C n'est pas moralement responsable du vol, car C n'était pas sensible aux raisons lorsqu'il a commis ce vol.

libre arbitre libertarianiste, nous sommes donc obligés de jongler avec deux visions diamétralement opposées de la responsabilité morale.

Avant de présenter la solution au problème présenté ci-haut, examinons les objectifs initiaux de Smilansky dans *Free Will and Illusion*. L'auteur s'intéressait essentiellement à trois questions : la question de la cohérence, la question de la compatibilité et la question des conséquences. La question de la cohérence se formule simplement : le libre arbitre libertarianiste est-il cohérent ? La réponse est négative. Cela nous amène donc à la question de la compatibilité : la responsabilité morale est-elle compatible avec l'absence de libre arbitre libertarianiste ? Le dualisme fondamental se présente comme la réponse à cette question. Selon Smilansky, il n'est pas raisonnable de répondre par « oui » ou « non » à la question de la compatibilité, il faut plutôt embrasser un compromis complexe entre ces deux réponses, d'où le dualisme fondamental qu'il propose. Cela nous amène maintenant à la question des conséquences : comment gérer l'absence de libre arbitre libertarianiste et le dualisme fondamental qui en découle ? C'est pour répondre à cette question que Smilansky développera son modèle illusionniste.

### **4.3 La solution : l'illusionnisme**

La solution de Smilansky, et celle que j'essaierai de défendre à partir de maintenant, est l'illusionnisme. Selon l'illusionnisme, le libre arbitre est une illusion. À quelle notion de libre arbitre fait-on référence ici ? Qu'entend-on par « illusion » ? Telles seront les questions auxquelles je tenterai de répondre dans cette section.

À quelle notion de libre arbitre avons-nous affaire ? Smilansky utilise une notion forte de libre arbitre. Plus précisément, il adopte une notion « robuste » de libre arbitre qu'il emprunte à certains libertarianistes.

A robust notion of libertarian free will would go much beyond compatibilism, and offer a large qualitative advantage in grounding desert, responsibility and justice. The most prominent current libertarianism, based upon event-libertarianism (e.g., Kane 1996; Ekstrom 2000) is not robust. It adds a measure of indeterminism to the initiation of action, which might be thought to limit and harm compatibilist control, but even when it does not, it does not add much to it. Robust libertarian free will is typically thought of as involving agent-causality. (Smilansky 2013, p. 106)

Smilansky croit donc que, dans le cadre d'une thèse illusionniste, la meilleure notion de libre arbitre est une notion libertarianiste et que la meilleure notion de contrôle est une notion de

contrôle *ultime* qui fait appel à la causalité de l'agent (voir chapitre 1). Pourquoi, en premier lieu, Smilansky emprunte-t-il une notion libertarienne de libre arbitre ? Tout simplement parce qu'une notion de libre arbitre qui est compatible avec le déterminisme est trop faible pour assoir toutes les attributions de responsabilité morale faites par les membres d'une communauté de responsabilité. Nous avons vu précédemment que les cas de manipulation échappaient aux compatibilistes, mais ce n'est pas tout. En effet, la faiblesse du compatibiliste est généralisée : « All that takes place on the compatibilist level, irrespective of the local distinctions in respect of control, becomes on the ultimate level "what was merely there", ultimately deriving from causes beyond the control of the participants. » (Smilansky 2000, p. 48) La plus intuitive des distinctions, comme celle faite dans l'exemple des triplets, devient peu convaincante lorsque l'on considère que personne ne peut choisir son caractère (conformément à la perspective valide des déterministes durs). Pourquoi choisir une notion libertarienne basée sur la causalité de l'agent ? Ce type de libertarisme met de l'avant une idée très séduisante, soit l'idée selon laquelle nous sommes capables de démarrer des chaînes causales à l'aide de notre volonté, sans être déterminé par quoi que ce soit. Cette idée, même si elle paraît peu satisfaisante à plusieurs, permet de justifier les attributions de responsabilité morale qui échappent au compatibilisme, contrairement au libertarisme qui fait seulement appel à la causalité des événements qui, selon Smilansky, n'apporte rien de bien significatif dans l'équation. L'important, pour le projet de Smilansky, n'est pas de mettre de l'avant une notion de libre arbitre qui est métaphysiquement correcte, mais bien de mettre de l'avant une notion de libre arbitre suffisamment forte pour dissiper le doute, aux yeux des membres d'une communauté de responsabilité, quant au caractère arbitraire des attributions de responsabilité morale. Examinons maintenant en quel sens le libre arbitre est une illusion.

Smilansky commence par noter qu'une condition minimale concernant la notion d'illusion est qu'il s'agit d'une croyance fautive (Smilansky 2000, p. 146). Une illusion est donc un certain type d'attitude propositionnelle où s'allie la fausseté d'un contenu propositionnel et le fait de tenir ce contenu pour vrai. Smilansky cherche ainsi à distinguer la notion d'illusion de celle de fiction. Une fiction est généralement conçue comme quelque chose de faux et qu'un agent sait être faux. Or, Smilansky ne cherche pas à soutenir que l'énoncé « j'ai un contrôle ultime sur mes actions » soit cru par un agent en dépit du fait que l'agent sait qu'il n'a pas de contrôle ultime sur ses actions.

Smilansky doit également donner une définition de la notion d'illusion qui permettra d'exclure les cas d'erreur puisqu'une erreur n'est pas une illusion. Smilansky écrit: « I use 'illusion' as a general term, focusing on various ways in which false beliefs are held, without complete awareness of their falseness, in the face of stronger epistemic claims to the contrary, which should have prevented them from being held. » (Smilansky 2000, p. 148) Grâce à cette condition, Smilansky réussit-il à exclure les cas d'erreur ? Il semble que oui. En effet, une des raisons pour lesquelles nous nous trompons est que nous ne sommes pas au courant qu'il existe des preuves contraires à notre croyance. Par exemple, il est possible de se tromper à propos des intentions écologistes du Parti Québécois après avoir lu, sans s'en rendre compte, un journal imprimé *avant* les élections. Dans ce cas, nous sommes confrontés à des preuves que le Parti Québécois a l'intention de lutter contre les gaz à effet de serre, et nous formons ainsi une croyance fautive par erreur. Dans le cas d'une illusion, contrairement aux cas d'erreur, il faut croire quelque chose de faux, et ce, même si nous sommes au courant qu'il existe des preuves qui peuvent contredire notre croyance.

De quelle manière est-il possible d'entretenir une croyance fautive en dépit de preuves contraires ? Smilansky discute principalement deux options : la pensée magique (*wishful thinking*) et la duperie de soi (*self-deception*)<sup>48</sup>. L'analyse de Smilansky, concernant ces deux notions, est très détaillée. Je me contenterai d'en retracer les grandes lignes. La pensée magique fait référence au fait de croire quelque chose parce qu'on a envie d'y croire. La pensée magique n'est donc pas fondée sur des données probantes ou sur un processus d'acquisition rationnel, elle est plutôt fondée sur le fait que certaines croyances sont plaisantes à entretenir. Certains énoncés, comme « je crois que le Canadien de Montréal gagnera la coupe Stanley cette année » ou « je crois que les vaches laitières ont une belle vie » exemplifient des cas de pensée magique. Selon Smilansky, l'illusion du libre arbitre pourrait être une instance de pensée magique. Par contre, à y regarder de plus près, la notion de pensée magique semble être trop faible en ce qui a trait au libre arbitre. En effet, dans les cas de pensée magique, un agent arrêtera d'entretenir sa croyance s'il est confronté à de fortes preuves du contraire (Smilansky 2000, p. 148). Un amateur de sport se rendra vite compte, en examinant les statistiques du Canadien de Montréal, que leurs chances de remporter la

---

<sup>48</sup> Smilansky (2000, p. 148) mentionne aussi l'aveuglement volontaire (*self-delusion*), mais il ne s'arrête pas sur cette notion.

coupe Stanley sont presque nulles. Un ami des animaux, s'il visite une ferme laitière, réalisera rapidement que les vaches laitières ont une vie misérable.

Dans le cas du libre arbitre, par contre, les individus semblent montrer beaucoup plus de résistance<sup>49</sup>. C'est pour cela, semble-t-il, que Smilansky préfère parler de duperie de soi. La duperie de soi fait référence au fait de croire quelque chose parce qu'on a envie d'y croire, et ce *en dépit de fortes preuves contraires*. Smilansky donne en exemple le cas d'un père de famille qui, malgré les allégations criminelles et les preuves démontrant la culpabilité de son fils, refuse de croire que celui-ci a commis le crime dont il est accusé (Smilansky 2000, p. 178). Semblablement, la croyance en un libre arbitre serait entretenue par certains agents en dépit de preuves contraires. Ainsi, certaines personnes pourraient croire qu'elles ont un contrôle ultime sur leurs actions, et ce même si elles sont au courant du paradigme déterministe en science et des découvertes sur le cerveau qui démontrent que nos processus cognitifs sont largement influencés par des facteurs génétiques et environnementaux. L'illusion de libre arbitre est ainsi considérée, par Smilansky, comme une instance de duperie de soi.

#### 4.4 Sauver la responsabilité morale

Ces précisions apportées sur le modèle illusionniste de Smilansky, nous pouvons nous demander s'il réussit bel et bien à sauver la responsabilité morale. La solution qui consiste à préférer l'illusion du libre arbitre à sa négation permet-elle de garantir la responsabilité morale de manière satisfaisante ? Divers arguments militent en ce sens. En effet, même si le libre arbitre n'existe pas, y croire paraît suffire à éviter les conséquences pratiques de son inexistence. Les agents, en croyant que leurs attributions de responsabilité morale sont justes, continueront à agir comme ils agissent présentement, c'est-à-dire à blâmer les mauvaises actions et à louer les bonnes, et surtout à se comporter eux-mêmes de manière à être louangés plutôt que blâmés. De plus, selon Smilansky, l'illusion du libre arbitre libertarienne est au fondement non seulement des comportements moraux, mais aussi de l'autovalorisation (Smilansky 2000, p. 170), de l'amitié, de la gratitude (p. 131), de l'amour et du respect (p. 176).

---

<sup>49</sup> Smilansky fonde cette affirmation sur plusieurs discussions qu'il a eues à propos du libre arbitre. Après plus de deux ans passés à défendre le scepticisme à propos du libre arbitre, je ne peux qu'être d'accord avec lui.

Les propos de Smilansky quant à l'importance de la croyance en un libre arbitre sur le plan psychologique et social, au moment de la publication de son livre, relevaient beaucoup plus de la spéculation qu'autre chose. Il est possible, maintenant, de défendre ses affirmations sur un plan empirique. Des études récentes tendent à montrer, par exemple, que l'absence de croyance en un libre arbitre favoriserait la tricherie (Vohs et Schooler 2008). Dans cette expérience, on demande à un groupe de participants de lire un texte qui renforce leur croyance en la vérité du déterminisme et on fait lire, à un autre groupe, un texte neutre. On demande ensuite aux participants de régler certains problèmes sur un ordinateur. Les participants ont l'opportunité de tricher pour régler ces problèmes. Il s'avère que le groupe ayant lu le texte pro-déterminisme triche significativement plus que le groupe ayant lu le texte neutre. Dans d'autres études, on a montré que la croyance en un libre arbitre encouragerait les comportements d'aide (Baumeister, Masicampo et DeWall 2009), favoriserait la gratitude (MacKenzie, Vohs et Baumeister 2014), et que son absence favoriserait les comportements agressifs (Baumeister, Masicampo et DeWall 2009). Le nombre d'expériences sur les bienfaits de la croyance en un libre arbitre ne cesse d'augmenter. En l'absence de théories satisfaisantes du libre arbitre, ces expériences militent en faveur de l'illusionnisme de Smilansky, dans la mesure où elles nous donnent des raisons de croire en l'importance de la croyance au libre arbitre. Il faudra attendre quelques années avant de connaître l'ampleur et l'importance que peut avoir la croyance au libre arbitre sur le comportement humain. Il semble, toutefois, qu'il faut pour l'instant donner raison à Smilansky lorsqu'il spéculait sur les effets bénéfiques de la croyance au libre arbitre au niveau moral, psychologique et social.

#### **4.5 Conclusion**

Nous avons finalement examiné le modèle sceptique illusionniste de Smilansky (2000). Selon ce modèle, le libre arbitre libertarien est une illusion. La nécessité d'une telle illusion naît d'une tension entre l'insuffisance du compatibilisme et la non-acceptabilité du déterminisme dur, dans un monde où le libre arbitre libertarien est impossible mais où la responsabilité morale doit être préservée. J'ai soutenu que le modèle de Smilansky réussit à éviter les problèmes rencontrés par le compatibilisme, le libertarienisme et le déterminisme dur. La notion de contrôle que Smilansky emprunte au libertarienisme capture très bien nos intuitions sur le libre arbitre, contrairement à celle proposée par les compatibilistes. L'illusionnisme de Smilansky évite les problèmes de nature ontologique du libertarienisme. Finalement, la position de Smilansky évite les problèmes de nature normative du détermi-

nisme dur en reconnaissant, preuves à l'appui, l'importance de la croyance en un libre arbitre et des attributions de responsabilité morale pour le maintien d'une société juste. J'espère ainsi avoir montré que le scepticisme à propos du libre arbitre peut être défendu sérieusement en philosophie.

## Conclusion

Avec le développement des sciences – et plus particulièrement des sciences physiques, des neurosciences et des sciences cognitives – il est devenu nécessaire de revisiter le problème du libre arbitre pour savoir si cette notion résiste aux découvertes récentes de la science. Est-il raisonnable d'affirmer l'existence d'un libre arbitre ? Intuitivement, il semble qu'il n'y ait pas de place pour un libre arbitre dans un monde déterministe, où le comportement humain est le résultat nécessaire des événements du passé en conjonction avec les lois de la nature. Même en admettant la fausseté du déterminisme, le seul indéterminisme admis en science se situe au niveau subatomique, et nous n'avons aucune raison de croire que l'indéterminisme quantique a des répercussions au niveau de l'action. Nous avons vu, toutefois, qu'il était possible de remettre en question les diverses menaces que soulèvent les sciences. Les compatibilistes, en développant des notions plus faibles de libre arbitre, prétendent conférer aux agents le contrôle nécessaire au libre arbitre (et à la responsabilité morale), et ce même si le déterminisme est vrai. Les libertarianistes, quant à eux, conservent une notion forte de libre arbitre et affirment du même souffle que le déterminisme est faux. La guerre entre le compatibilisme et le libertarianisme a fait, et continue de faire, couler beaucoup d'encre. Il existe pourtant une autre voie, celle du scepticisme.

Dans le premier chapitre de ce mémoire, l'objectif était de sélectionner, clarifier et exemplifier les diverses approches qu'offre la littérature philosophique contemporaine pour une défense du libre arbitre. J'ai commencé par distinguer le libre arbitre de la liberté et de la responsabilité morale, pour ensuite examiner deux visions de la causalité dans le débat sur le libre arbitre. Par la suite, j'ai évalué deux notions communes de libre arbitre : la notion contre-causale et la notion faisant appel à l'initiation consciente de nos actions. Ces deux derniers modèles furent rejetés, le premier parce qu'il a recours à une causalité de l'agent, et le deuxième parce que certaines expériences en neuroscience rendent ce modèle peu plausible. J'ai ensuite exposé l'un des modèles les plus respectés de libertarianisme, le modèle indéterministe de Laura Ekstrom. Pour Ekstrom, une action sera dite « libre » si elle a été causalement déterminée par une intention qui a été causalement déterminée par une préférence qui a elle-même été causée, de manière indéterminée, par des considérations lors d'une délibération. Ekstrom cherchait ainsi à mettre de l'avant que le libre arbitre exige une certaine part d'indétermination dans la formation de notre caractère. J'ai ensuite tourné mon attention vers deux modèles compatibilistes de libre arbitre. Le premier, celui d'Harry Frankfurt, mise

sur les notions de désirs du premier ordre et de désir du second ordre. Une action sera dite « libre » si elle a été produite par un désir du premier ordre qui est conforme à un désir effectif du second ordre. Le second, celui de John Martin Fischer et Mark Ravizza, lie le libre arbitre à une sensibilité aux raisons. Cette notion contrefactuelle de libre arbitre met en évidence qu'une action, pour être libre, doit être une réponse à des considérations rationnelles. Je crois, dans ce premier chapitre, avoir jeté les bases d'une investigation sérieuse sur le libre arbitre.

Dans le deuxième chapitre de ce mémoire, j'ai entrepris une critique du compatibilisme. J'ai commencé par exposer et défendre l'argument de la conséquence de Peter van Inwagen. Cet argument nous dit que si le déterminisme est vrai, alors notre manque de contrôle sur les événements du passé et sur les lois naturelles engendre un manque de contrôle sur nos actions. Grâce à l'argument de la conséquence, j'ai soutenu que le libre arbitre n'est pas compatible avec le déterminisme, contrairement aux prétentions des compatibilistes. Par la suite, j'ai exposé l'argument de la manipulation en quatre étapes, formulé par Derk Pereboom. Pereboom lance deux défis aux compatibilistes : rendre compte de l'absence de libre arbitre dans un exemple sophistiqué de manipulation et distinguer ce cas de manipulation des cas de déterminisme « ordinaire » qui prétendument serait compatible avec le libre arbitre. Après avoir posé une distinction entre les modèles historiques et non historiques de libre arbitre, j'ai soutenu que ces deux types de modèle échouent à la première étape du défi lancé par Pereboom. En limitant mon évaluation au modèle de Frankfurt et au modèle de Fischer et Ravizza, je ne prétends pas avoir miné totalement la crédibilité des modèles compatibilistes. Je crois par contre avoir donné de bonnes raisons de douter de la plausibilité du projet compatibiliste (concilier le libre arbitre et le déterminisme).

Dans le troisième chapitre de ce mémoire, j'ai entrepris une critique du libertarianisme. J'ai commencé par évaluer l'argument de la chance de van Inwagen. Selon cet argument, l'indéterminisme des modèles libertarianistes engendre de la chance, de telle sorte qu'il rend impossible toute action libre. Ayant été convaincu par la réponse qu'Ekstrom adresse à van Inwagen, j'ai jugé que l'argument de la chance était invalide, et je me suis tourné vers un autre argument, inspiré de l'argument de base de Galen Strawson. Grâce à cet argument, j'ai soutenu que l'indéterminisme des modèles libertarianistes ne permet pas d'attribuer le contrôle qu'un agent doit avoir pour être qualifié de libre. Par la suite, j'ai exposé le quasi-déterminisme de Ted Honderich. Selon Honderich, nous sommes entièrement dé-

terminés au niveau de notre existence. Comme nous l'avons vu, les libertariens ont besoin d'un moment pertinent d'indétermination dans la chaîne causale des actions, et si le déterminisme est vrai, il semble que ces moments pertinents d'indétermination n'existent tout simplement pas. Encore une fois, je ne prétends pas, en ayant évalué un seul modèle libertarien, avoir miné complètement la crédibilité des modèles libertariens. Je crois par contre avoir donné de bonnes raisons de douter de la plausibilité du projet libertarien (concilier le libre arbitre et l'indéterminisme et mettre en doute la vérité du déterminisme).

Dans le quatrième chapitre de ce mémoire, j'ai exposé et défendu le modèle illusionniste de Smilansky. Ce modèle sceptique met de l'avant l'idée selon laquelle le libre arbitre libertarien est une illusion, c'est-à-dire une croyance fautive tenue en dépit de preuves contraires. Selon Smilansky, l'illusion du libre arbitre libertarien est normativement nécessaire au maintien de la société, et il ne faut pas chercher à s'en débarrasser. J'ai soutenu que le modèle de Smilansky est plus plausible que les modèles concurrents (le compatibilisme, le libertarisme et le déterminisme dur). Je crois ainsi qu'il est préférable d'adopter une position sceptique, et plus particulièrement illusionniste, à propos du libre arbitre.

Mon objectif était de paver la voie à une défense sérieuse du scepticisme à propos du libre arbitre. Selon moi, la littérature contemporaine sur le libre arbitre est trop souvent dominée par les approches compatibiliste et libertarien, et on oublie qu'il existe une manière de s'extirper de cette dichotomie. Après réflexion, il me semble que le libre arbitre n'est ni compatible avec le déterminisme ni compatible avec l'indéterminisme. C'est pourquoi je considère qu'il faut adopter un scepticisme à propos du libre arbitre, et j'espère avoir réussi à donner de la crédibilité à cette thèse.

Dans ce mémoire, j'ai surtout souligné que le scepticisme échappait aux problèmes traditionnels et qu'il représentait une alternative prometteuse au libertarisme et au compatibilisme. Par contre, je me suis peu intéressé aux débats et aux problématiques proprement sceptiques. Par exemple, une fois que l'on a admis que le libre arbitre n'existe pas, sommes-nous contraints de vivre « comme si » on était libre ? En d'autres mots, la croyance en un libre arbitre est-elle une nécessité psychologique ? D'un côté, Daniel Wegner (2002) croit qu'il est impossible, dans la vie de tous les jours, d'agir sans se croire libre. La plupart des philosophes, par contre, considèrent qu'il est possible d'échapper à cette croyance (Blackmore 2013). Admettons maintenant, pour le bien de la discussion, qu'il est possible

d'échapper à la croyance en un libre arbitre. Une fois que cela est admis, il est légitime de se demander : *devrait-on* se débarrasser de cette croyance ? S'agit-il de la meilleure chose à faire ? D'un côté, les préservationnistes croient qu'il est préférable de conserver l'illusion du libre arbitre (Smilansky 2000). Smilansky, comme nous l'avons vu, considère que l'illusion du libre arbitre est nécessaire au maintien de notre société. L'illusionnisme de Smilansky se veut ainsi une forme non radicale de scepticisme. Certains croient plutôt qu'il serait préférable d'abandonner cette illusion et adoptent ainsi une forme plus radicale de scepticisme (Pereboom 2001; Strawson 1986; Blackmore 2013; Waller 2013). Ces éliminativistes à propos du libre arbitre soutiennent par exemple que le libre arbitre n'est pas, tout compte fait, une condition nécessaire à la responsabilité morale ou que les impacts positifs associés à l'abandon de la croyance en un libre arbitre ou à l'abandon de la responsabilité morale surpassent les impacts négatifs. Il serait intéressant, dans un futur travail, d'explorer plus profondément ces questions.

## Bibliographie

- Anscombe, G. E. M. (1993). "Causality and Determination". *Causation*, E. Tooley et M. Sosa (dir.), Oxford: Oxford University Press, 88-104.
- Aristote, (*EN*) *Éthique à Nicomaque*. trad. Jean Tricot, Paris: Vrin, 1959.
- Berlin, Isaiah (1969). "Two Concepts of Liberty". *Four Essays on Liberty*, I. Berlin (dir.), Oxford: Oxford University Press.
- Bishop, Robert H. (2009). "Chaos". *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (dir.).
- (2011). "Chaos, Indeterminism, and Free Will". Kane, Robert (dir.) (2011), 84-100.
- Blackmore, Susan. (2013). "Living Without Free Will". Caruso, Gregg D. (dir.) (2013), 161-175.
- Brembs, Björn (2011). "Towards A Scientific Concept Of Free Will As A Biological Trait: Spontaneous Actions And Decision-Making In Invertebrates". *Biological Sciences*, 278, 930-939.
- Carter, Ian (2012). "Positive and Negative Liberty". *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (dir.).
- Campbell, C.A. (1967). *In Defense of Free Will with Other Philosophical Essays*, London: George Allen et Unwin.
- Caruso, Gregg D. (dir.) (2013). *Exploring the Illusion of Free Will and Moral Responsibility*, New-York: Lexington Books.
- Casati, Roberto et Varzi, Achille (2014). "Events". *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (dir.).
- Chisholm, Roderick (1976). "The Agent as Cause". *Action Theory*, M. Brand et D. Walton (dir.), Reidel, 199-211.
- (1982). "Human Freedom and the Self". *Free Will*, Gary Watson (dir.), Oxford: Oxford University Press, 24-35.
- Dennett, Daniel (2013). "A Dialogue on Free Will". *Method. Free Will: Thirty points of views*, 2, 3, 46-48.
- Earman, John (1986). *A Primer on Determinism*, New-York: Springer Science & Business Media.
- Eshleman, Andrew (2014). "Moral Responsibility". *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (dir.).
- Ekstrom, Laura Waddell (2000). *Free Will: A Philosophical Study*, Westview Press.

- (2011). “Free Will is not a Mystery”. Kane, Robert (dir.) (2011), 366-380.
- Finch, Alicia et Warfield, Ted (1998). “The Mind Argument and Libertarianism”. *Mind*, 107, 427, 515-528.
- Fischer, John Martin (1995). *The Metaphysics of Free Will: An Essay on Control*, Cambridge: Wiley Blackwell.
- (2004). “Responsibility and Manipulation”. *Journal of Ethics*, 8, 2, 145-177.
- (2006). *My Way: Essays on Moral Responsibility*, Oxford: Oxford University Press.
- (2007). “Response to Kane, Pereboom and Vargas”. Fischer, John Martin et Al. (dir.) (2007), 184-190.
- Fischer, John Martin et Ravizza, Mark (1998). *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.
- Fischer, John Martin et al. (dir.) (2007). *Four Views on Free Will*, Oxford: Wiley Blackwell.
- Foley, Richard (1979). “Compatibilism and Control over the Past”. *Analysis*, 39, 2, 70-74.
- Frankfurt, Harry (1969). “Alternative Possibilities and Moral Responsibility”. *Journal of Philosophy*, 66, 3, 829-839.
- (1971). “Freedom of the Will and the Concept of a Person”. *Journal of Philosophy*, 68, 1, 5-20.
- (2002). “Reply to John Martin Fischer”. *Contours of Agency: Essays on Themes from Harry Frankfurt*, S. Buss et L. Overton (dir.), Cambridge: MIT Press, 27-32.
- Garson, James W. (1995). “Chaos and Free Will”. *Philosophical Psychology*, 8, 4, 365-374.
- Griffith, Meghan (2007). “Freedom and Trying: Understanding Agent-Causal Exertions”. *Acta Analytica*, 22, 1, 16-28.
- Hobbs, Jesse (1991). “Chaos and Indeterminism”. *Canadian Journal of Philosophy*, 21, 2, 141-164.
- Honderich, Ted (2002). *How Free Are You? : The Determinism Problem*, Oxford: Oxford University Press.
- (2013). “Determinism, Incompatibilism and Compatibilism”. Caruso, Gregg D. (dir.) (2013), 53-64.
- Hoefer, Carl (2010). “Causal Determinism”. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (dir.).
- Hume, David (1738). *A Treatise of Human Nature*. London: Penguin Classics, 1985.
- Kane, Robert (dir.) (1996). *The Significance of Free Will*, Oxford: Oxford University Press.

- (2011). *The Oxford Handbook of Free Will: Second Edition*, Oxford: Oxford University Press.
- Libet, Benjamin (1999). “Do We Have Free Will”. *Journal of Consciousness Studies*, 6, 8-9, 47-57.
- MacKenzie, Michael J., Vohs, Kathleen D. et Baumeister, Roy F. (2014). “You Didn't Have to Do That: Belief in Free Promotes Gratitude”, *Personality and Social Psychology Bulletin*, 40, 11, 1423-1434.
- Mawson, Tim (2011). *Free Will: A Guide for the Perplexed*, London: Bloomsbury Academic.
- McFee, Graham (2001). *Free Will*, Montreal: McGill-Queens University Press.
- McKay, Thomas et Johnson, David (1996). “A Reconsideration of an Argument Against Compatibilism”. *Philosophical Topics*, 24, 2, 113-122.
- McKenna, Michael et D. Justin Coates (2009). “Compatibilism”. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (dir.).
- Mele, Alfred (1995). *Autonomous Agents: From Self-Control to Autonomy*, Oxford: Oxford University Press.
- (2005). “A critique of Pereboom's ‘Four-Case Argument’ for Incompatibilism”. *Analysis*, 65, 285, 75-80.
- (2007). “Free Will: Action Theory Meets Neuroscience”. *Intentionality, Deliberation, and Autonomy: The Action-Theoretic Basis of Practical Philosophy*, Lumer C. (dir.), Ashgate, 257-272.
- (2011). “Libet on Free Will: Readiness Potentials, Decisions, and Awareness”. *Conscious Will and Responsibility*, W. Sinnott-Armstrong et L. Nadel (dir.), Oxford: Oxford University Press, 23-33.
- Nadelhoffer, Thomas et Goya Tocchetto, Daniela (2013). “The Potential Dark Side of Believing in Free Will (and Related Concepts): Some Preliminary Findings”. Caruso, Gregg D. (dir.) (2013), 121-140.
- O'Connor, Timothy (1995). “Agent Causation”. *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, T. O'Connor (dir.), Oxford: Oxford University Press, 61-79.
- (2005). “Freedom With a Human Face”. *Midwest Studies in Philosophy*, 29, 1, 207-227.
- (2014). “Free Will”. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (dir.).
- Pereboom, Derk (2001). *Living Without Free Will*, Cambridge University Press.
- (2007a). “Hard Incompatibilism”. Fischer, John Martin et Al. (dir.) (2007), 85-125.
- (2007b). “Response to Kane, Fischer and Vargas”. Fischer, John Martin et Al. (dir.) (2007), 191-203.

- (2013). "Skepticism about Free Will". Caruso, Gregg D. (dir.) (2013), 19-40.
- Pockett, Susan (2013). "If Free Will Did Not Exist, It Would Be Necessary to Invent It". Caruso, Gregg D. (dir.) (2013), 265-272.
- Rowe, William (2006). "Free Will, Moral Responsibility, and the Problem of 'OOMPH'". *The Journal of Ethics*, 10, 3, 295-313.
- Slote, Michael (1982). "Selective Necessity and the Free Will Problem". *Journal of Philosophy*, 79, 5-24.
- Smilansky, Saul (2000). *Free Will and Illusion*, Oxford: Oxford University Press.
- (2013). "Free Will as a Case of "Crazy Ethics"". Caruso, Gregg D. (dir.) (2013), 103-120.
- Sorabji, Richard (1980). *Necessity, Cause and Blame*, New-York: Cornell University Press.
- Strawson Galen (1986). *Freedom and Belief*. Oxford: Oxford University Press.
- (2013). "The Impossibility of Ultimate Responsibility". Caruso, Gregg D. (dir.) (2013), 41-52.
- Steward, Helen (2012). *A Metaphysics for Freedom*, Oxford: Oxford University Press.
- Taylor, Richard (1991). *Metaphysics, 4th Edition*, Englewood Cliffs, N.J: Prentice Hall.
- Van Inwagen, Peter (1986). *An Essay on Free Will*, Oxford: Oxford University Press.
- (2000). "Free Will Remains a Mystery: The Eighth Philosophical Perspectives Lecture". *Noûs*, 34, 1-19.
- Vihvelin, Kadri (2011). "Arguments for Incompatibilism". *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (dir.).
- Vohs, Kathleen D. et Schooler, Jonathan W. (2008). "The Value Of Believing In Free Will: Encouraging A Belief In Determinism Increases Cheating", *Psychological Science*, 19, 1, 49-54.
- Waller, Bruce (2013). "The Stubborn Illusion of Moral Responsibility". Caruso, Gregg D. (dir.) (2013), 65-86.
- Wallace, R. Jay (1994). *Responsibility and the Moral Sentiments*, Boston: Harvard University Press.
- Wegner, Daniel (2002). *The Illusion of Conscious Will*, Cambridge: The MIT Press.
- Wiggins, David (1970). "Freedom, Knowledge, Belief and Causality". *Royal Institute of Philosophy Lectures*, 3, 132-154.