

Université de Genève
Département de Philosophie

ATTITUDES RÉACTIVES ET DÉTERMINISME

Mémoire

Par Adrian Sulja

Directeur : Prof. Julien Deonna

Remerciements

Je tiens à remercier, en tout premier lieu, le professeur M. Julien Deonna, directeur de ce mémoire, qui, par sa relecture précise, ses commentaires avisés et nos riches échanges, m'a permis de mener à bien ce travail.

Je remercie, ensuite, le professeur assistant M. Simon-Pierre Chevarie-Cossette, ainsi que les doctorants M^{me} Agnès Baehni, M. Robin Bianchi et M. Antoine Rebours qui m'ont apporté leurs expertises et conseils dans les différents points clefs de mes réflexions.

J'aimerais remercier également, mon camarade d'université et ami M. Didier Kofmehl pour sa relecture, ainsi que nos précieuses discussions philosophiques qui n'ont jamais cessées pendant nos années de formation.

Enfin, je remercie mes proches pour le soutien et l'amour qu'ils m'apportent et, si cela est raisonnable, toutes les causes silencieuses qui ont rendu ce travail nécessaire.

Table des Matières

❖ INTRODUCTION	3
❖ CHAPITRE I : ATTITUDES REACTIVES ET RESPONSABILITE MORALE	4
1.1 Attitudes Réactives ou Emotions Morales.....	5
1.2 Deux Approches de la Responsabilité Morale	12
1.3 Thèse Traditionnelle de la Responsabilité Morale	14
❖ CHAPITRE II : SCEPTICISME A PROPOS DES THESES TRADITIONNELLES DE LA RESPONSABILITE MORALE.	18
2.1 Déterminisme et Incompatibilisme Primitif	19
2.2 Le Problème du Libre Arbitre.....	23
2.3 Compatibilisme	24
2.3.1 <i>La Thèse du Désir de Haut Niveau : Frankfurt</i>	27
2.3.2 <i>La Thèse de la Sensibilité aux Raisons : Fischer et Ravizza</i>	28
2.4 Arguments contre le Compatibilisme	31
2.4.1 <i>Chance Constitutive</i>	31
2.4.2 <i>Argument de la Manipulation</i>	34
2.5 Libertarianisme	41
2.5.1 <i>Libertarianisme de la Causalité Agentive</i>	42
2.5.2 <i>Libertarianisme de la Causalité Evènementielle</i>	43
2.6 Arguments contre le Libertarianisme	45
2.6.1 <i>Causa Sui</i>	45
2.6.2 <i>Hasard</i>	47
2.7 Du Scepticisme au Déterminisme Dur	49
❖ CHAPITRE III : LE PROBLEME DES ATTITUDES REACTIVES	51
3.1 Identification des Problèmes du Déterminisme Dur	52
3.1.1 <i>Moralité</i>	54
3.1.2 <i>Le Sens de la Vie</i>	57
3.1.3 <i>Les Attitudes Réactives</i>	58
3.2 Freedom and Resentment	62
3.2.1 <i>Thèse Naturaliste</i>	64
3.2.2 <i>Naturalisme et Responsabilité Morale</i>	66
3.2.3 <i>Thèse Normative</i>	68
3.2.4 <i>Comment Strawson répond-il aux problèmes des sceptiques ?</i>	69
3.3 Critiques de la thèse naturaliste	71
3.4 Critiques de la Thèse Normative.....	76
3.5 Vers L'Illusionnisme	80
❖ CONCLUSION	83
❖ BIBLIOGRAPHIE	84

Introduction

En philosophie contemporaine, l'une des questions métaphysiques les plus centrales concerne la compatibilité entre déterminisme et libre arbitre. Le compatibilisme a connu, à la suite des travaux précurseurs de David Hume, un véritable succès philosophique que l'on doit aux développements ultérieurs de certains philosophes du XXème siècles, tels que Harry Frankfurt et Peter Frederick Strawson. Depuis quelques décennies, l'incompatibilisme effectue néanmoins un retour en force qui remet d'anciennes questions au goût du jour. Dans ce travail, je m'intéresse précisément à l'un de ces problèmes que les strawsoniens croyaient avoir résolu : les attitudes réactives sont-elles compatibles avec une croyance au déterminisme ? Pour répondre à cette question, je reprends d'abord la notion d'attitudes réactives proposée par Strawson afin de la redéfinir et conçois les attitudes réactives comme des émotions qui impliquent une attribution de responsabilité morale. Je m'intéresse, ensuite, au débat entre déterminisme et libre arbitre et prends parti pour le déterminisme dur.

Cette position me mène alors à un problème : 1) si nos attitudes réactives impliquent une attribution de responsabilité morale et 2) si la responsabilité morale est impossible, alors nos attitudes réactives sont toujours inappropriées, parce qu'elles représentent une responsabilité morale qui n'existe pas. Devons-nous, dès lors, abandonner l'ensemble de ces pratiques qui sont pourtant si quotidiennes et fondamentales ? Bien que la solution de Strawson ne me convainque pas pleinement, je crois, comme lui, qu'il est dans l'intérêt des déterministes de trouver le moyen de sauver ces pratiques.

Chapitre I : Attitudes Réactives et Responsabilité Morale

Dans ce premier chapitre, je m'intéresse au lien particulier qu'entretiennent certaines émotions avec la responsabilité morale. Celles-ci ont reçu une attention pionnière à l'occasion d'un fameux article de P.F. Strawson¹ qui, le premier, a attiré le débat du libre arbitre sur la question des attitudes réactives. Pour Strawson, les attitudes réactives désignent un ensemble d'attitudes qui réagissent aux bonnes ou mauvaises volontés qui sont exprimées par les participants d'une communauté morale donnée. Considérons la situation suivante : je suis dans un bus bondé et quelqu'un me marche sur le pied. Si la personne avait l'intention claire de me faire du mal, alors il est probable que je ressente de la colère envers elle. En revanche, si cette personne n'exprime pas d'intention malveillante et me blesse par accident, alors il est peu probable que je ressente de la colère envers elle. La colère est donc un bon exemple d'attitude réactive. En effet, cette émotion montre que la nature de la volonté manifestée par les personnes qui nous entourent (bienveillante, malveillante ou indifférente) nous importe beaucoup et détermine nos réactions.

Aussi, le but de ce chapitre est l'introduction d'un certain nombre de réflexions générales sur le sujet. Néanmoins, je me sers de la notion strawsonienne d'attitudes réactives, non pas pour défendre les thèses strawsonniennes qui seront présentées plus tard, mais seulement parce qu'elles me semblent être un bon point de départ. À ce stade préliminaire, je pars simplement de l'idée que les attitudes réactives existent et essaie de les analyser pour elles-mêmes, en dehors de la fonction qu'elles occupent dans les réflexions de Strawson.

Dans un premier temps, je présenterai les attitudes réactives et tenterai de les définir comme des émotions morales, en mettant en lumière le lien qu'elles entretiennent avec la responsabilité (**1.1.**). Dans un second temps, je tâcherai de définir la notion de responsabilité morale qui est au cœur du problème (**1.2.**). J'étudierai, à cet effet, la définition traditionnelle de la responsabilité morale (**2.3.**).

¹ Strawson, P. F. (1962). *Freedom and resentment*. Proceedings of the British Academy, 48, 1-25.

1.1 Attitudes Réactives ou Emotions Morales

Les attitudes réactives² ont intégré le lexique philosophique contemporain, à l'occasion de la publication, désormais fameuse, de *Freedom and Resentment*, de Peter Strawson. Elles désignent un ensemble de réactions émotionnelles humaines qui se manifestent à l'encontre des intentions que nous percevons chez les autres personnes (et nous-même), dans la mesure où celles-ci sont exprimées dans nos comportements et actions. Certaines de nos émotions semblent être particulièrement sensibles aux bonnes ou mauvaises volontés (good or ill will)³ qui apparaissent dans un contexte social. Nous nous mettons souvent en colère, par exemple, dès lors que nous nous trouvons offensés et ressentons de la gratitude, à l'inverse, lorsque nous croyons être particulièrement considérés et aimés. Les attitudes réactives désignent, donc, un ensemble relativement large d'attitudes qui ont pour caractéristique commune d'être autant de réactions aux bonnes et mauvaises volontés exprimées socialement. Ces attitudes sont le plus souvent des émotions, comme la colère, le ressentiment, la gratitude, l'indignation, l'approbation, la honte, le remord, la culpabilité, ainsi que certaines formes d'amour et de fierté. Mais d'autres attitudes comme le pardon et le blâme peuvent également compter parmi les attitudes réactives.

En outre, Strawson propose de distinguer trois catégories d'attitudes réactives, en vertu de leur orientation. Il y a, premièrement, les attitudes réactives personnelles⁴ qui reflètent **une attente dirigée vers les autres**. Ce type de réactions exprime essentiellement la demande, à la première personne, qu'une exigence soit satisfaite. Ces exigences portent par exemple sur le fait de ne pas être offensé, de ne pas être humilié ou plus généralement de recevoir un degré suffisant de bonnes intentions. La colère reflète, par exemple, la demande d'être respecté à sa juste valeur, laquelle, si elle est insatisfaite, cause la colère.

Il y a, deuxièmement, les attitudes réactives réflexives⁵ qui reflètent **une attente dirigée vers soi-même**. La honte, le remord et la culpabilité sont des émotions qui reflètent une exigence dirigée vers soi-même, en particulier celle de ne pas transgresser de normes morales ou plus généralement de ne pas faillir trop gravement aux valeurs auxquelles l'on est attaché.

² De l'anglais « Reactive attitudes ».

³ Strawson, P. F. (1962). *Freedom and resentment*. Proceedings of the British Academy, 48, 1-25.

“The reactive attitudes I have so far discussed are essentially reactions to the quality of others' wills towards us, as manifested in their behaviour: to their **good or ill will or indifference or lack of concern**.”

⁴ Ibid. « personal reactive attitudes ».

⁵ Ibid. « self-reactive attitudes ».

Troisièmement, il y a les attitudes réactives impersonnelles ou indirectes⁶ qui sont une généralisation des attitudes réactives personnelles. Elles manifestent l'attente générale que certaines exigences soient satisfaites non seulement pour soi, mais dans un contexte interpersonnel très large. On peut donc les concevoir comme reflétant une **attente dirigée vers les autres et pour les autres**. En effet, certaines réactions émotionnelles, telles que l'indignation et l'approbation, ne semblent pas se limiter à une attente personnelle. Nous sommes parfois indignés par des actions qui ne nous concernent pas directement et qui se déroulent à des milliers de kilomètres. Ces attitudes reflètent, ainsi, une attente dirigée vers les êtres humains et pour les êtres humains.

Cependant, on peut se demander s'il existe véritablement une unité entre ces différents types d'émotions. Sur quelle base commune les attitudes réactives réagissent-elles aux bonnes et mauvaises volontés que nous identifions ?

Une bonne façon de les caractériser consiste, je pense, à concevoir les attitudes réactives comme des « émotions morales ». Mais il convient d'écartier au plus vite une confusion qui peut naître de la polysémie de ce terme.

Lorsque je dis que les attitudes réactives peuvent être conçues comme des émotions morales, je ne veux pas supposer qu'il y aurait des émotions systématiquement morales d'une part, et immorales d'autre part. On a coutume de juger, par exemple, que la jalousie est toujours immorale et l'empathie toujours morale⁷, dans la mesure où « moral » et « immoral » sont traités ici comme des synonymes de « bon » et « mauvais ».

Cependant, le sens de « moral » qui est pertinent dans notre cas, est plutôt celui qui s'oppose à « amoral ». Les attitudes réactives pourraient être conçues comme morales, non pas parce qu'elles seraient de « bonnes » émotions, mais parce que leur justification fait toujours appel à des concepts moraux. En effet, les attitudes réactives sont, en premier lieu, des attitudes intentionnelles qui portent sur un objet précis. Nous sommes indignés *de* ou honteux *de* X. À ce titre, affirmer qu'elles sont des émotions morales consiste seulement à reconnaître que leurs objets formels sont typiquement issus de notre registre moral. L'offense et la faute sont des notions qui peuvent être analysées en termes moraux, car elles semblent impliquer la croyance qu'une norme morale a été transgessée. Les attitudes réactives sont donc dites « morales » car elles représentent les intentions exprimées dans nos comportements selon un critère nécessairement moral. Elles semblent, en effet, entretenir un rapport privilégié avec les valeurs morales plutôt que les autres familles de valeurs. Ainsi, la culpabilité d'avoir volé de l'argent est une émotion morale, parce que mon intention est représentée comme une faute morale.

⁶ Ibid. « vicarious reactive attitudes ».

⁷ On peut certainement invoquer ici un héritage judéo-chrétien.

Selon cette caractérisation, la peur ne peut être comprise comme une émotion morale, puisque l'objet formel qu'elle capture typiquement, à savoir le danger, n'a aucune valeur morale spécifique.

Mais est-il vrai que les attitudes réactives font nécessairement appel à des critères moraux ? Il me semble, de prime abord, qu'il doit être possible de trouver des exemples de culpabilité ou de colère qui n'impliquent aucun contenu moral. Une personne de confession musulmane pourrait, par exemple, ressentir de la culpabilité après avoir bu de l'alcool, sans pour autant supposer que la norme qui a été transgressée soit de nature morale, mais plutôt de nature religieuse.

De la même manière, il paraît possible de se sentir coupable de n'avoir pas fait assez de sport, d'avoir échoué à un examen ou d'avoir dépensé trop d'argent pour des raisons qui ne sont pas morales. Plus largement, nous nous mettons souvent en colère pour un ensemble assez large de raisons qui ne supposent pas de manquements moraux évidents. Dans tous ces cas, je ne suis pas sûr que ces émotions représentent la transgression sur un plan nécessairement moral, mais relativement à d'autres types d'obligations (religieuses, éducationnelles, conventionnelles etc.). En bref, cette conception des attitudes réactives comme émotions qui représentent les objets selon un critère morale semble d'emblée compromise, car elle est trop restreinte.

Une façon prometteuse d'envisager les attitudes réactives comme des émotions morales consiste désormais à mettre l'accent sur leur croyance d'arrière-plan. Toutes les attitudes réactives semblent impliquer, en effet, une croyance qui porte sur la responsabilité morale. Mais cette croyance ne doit toutefois pas être confondue avec celle que nous venons de rejeter. Lorsque nous nous mettons en colère, nous croyons que la personne visée est responsable, sans nécessairement croire qu'elle est moralement fautive.⁸ Croire qu'une personne est responsable n'implique donc pas que celle-ci a nécessairement transgressé une règle morale, mais seulement qu'elle incarne un type d'agent moral susceptible d'être blâmé ou loué ou, en d'autres termes, qui est capable de *répondre* de ses actes et de *mérirer* les blâmes et louanges. Ainsi, en suivant cette idée, la responsabilité peut être conçue comme une capacité morale, se traduisant par un contrôle suffisant sur nos actions pour mérirer d'être blâmé ou loué.

⁸ Pour reprendre mon exemple, il y a un sens selon lequel je suis responsable de ne pas avoir fait assez de sport, sans avoir enfreint de règles morales pour autant.

Ce contrôle est traditionnellement assimilé, selon bon nombre de philosophes au principe des possibilités alternatives :

"Strawson reminds us how often in our relations with one another we display and think that we have good reason to display such feelings as gratitude and resentment, how we take pride in our achievements and feel remorse for our misdeeds, how actions to which we are not a party can evoke our moral approval or indignation. In all these cases there is a tacit or explicit reference to the concept of desert. A person who does me a benefit deserves my gratitude because he could have withheld the favour. I feel remorse for my neglect of my friends, because I could have taken more trouble to please them. In my hatred of tyranny, wherever it is displayed, the idea is implicit that the tyrant could have restrained himself, or even if this is no longer possible, that he himself is at least partly to blame for having become the monster that he is."⁹

Ainsi, dire que les attitudes réactives sont des émotions morales revient, finalement, à souligner le lien qu'elles entretiennent avec la responsabilité définie en termes de contrôle.¹⁰ Il est à noter que ce lien particulier n'apparaît pas à l'analyse des autres émotions. Il constitue, de ce fait, la particularité propre aux attitudes réactives. En effet, si j'ai peur de mon oncle ou si je suis en deuil suite au décès d'un ami, la croyance que ces deux personnes avaient ou non la capacité d'agir autrement ne semble en aucun cas pertinente pour l'émotion qui est vécue. Invariablement, ces émotions continuent à avoir du sens, peu importe le contrôle agentiel.

Or, dans le cadre des attitudes réactives, cette base cognitive apparaît comme essentielle et constitutive de l'émotion. De la même manière qu'il est nécessaire, pour avoir peur d'une araignée, de se la représenter comme présente (typiquement en la voyant, l'imaginant, en s'en souvenant, etc...), il est nécessaire aux attitudes réactives de représenter la personne visée par l'émotion comme responsable. Certains exemples d'attitudes réactives montrent relativement bien cette nécessité. La fierté d'avoir secouru un enfant au bord de la noyade, l'indignation face à la tyrannie d'un régime politique ou la culpabilité à la suite d'une mauvaise prise de décision sont des réactions émotionnelles qui représente l'action comme produite librement. Sans cette croyance d'arrière-plan, ces émotions paraissent irrationnelles. Considérons à l'inverse : la fierté d'avoir un joli visage, l'indignation face au comportement d'un nouveau-né ou la honte d'être un homme nous paraissent effectivement irrationnelles, précisément parce qu'elles représentent une capacité qui fait défaut. De la même manière, se mettre en colère contre une voiture en panne est irrationnel, parce que l'attitude représente une propriété qui fait défaut à la voiture.

⁹Ayer, A. J. (1980). *Free Will and Rationality*. In van Straaten, R. (Ed.), *Philosophical Subjects* (pp. 211-224). University of California Press

¹⁰ Nous verrons, à la suite de ce travail, quelles sont précisément les conditions de la responsabilité morale.

Ainsi, même si certaines de nos occurrences d'attitudes réactives devaient avoir un caractère irrationnel, ce qui n'est pas fondamentalement impossible, il n'en demeure pas moins qu'une condition de responsabilité est nécessaire à la rationalité de l'émotion.

Ce caractère nécessaire est d'ailleurs nettement perçu dans certains cas de révision ou de rationalisation de nos émotions. Par exemple, il paraît approprié d'être en colère contre mon voisin, si celui-ci est entré par effraction dans mon appartement pendant la nuit. Mais cette émotion sera volontiers soumise à révision si j'apprends, dans un second temps, que ce voisin était victime d'une crise de somnambulisme au moment des faits. Dans ce cas, c'est précisément parce qu'il n'est plus possible de se représenter le voisin comme responsable, que la colère disparaît.

Pour plus de clarté, prenons les cas de la colère et du remord qui sont particulièrement bien décrits chez Roberts (2003), lequel cherche à individuer les émotions par le biais des familles de valeurs. L'offense et la faute sont des valeurs qui ont pour point commun de faire effectivement appel à la responsabilité. L'ensemble des émotions voisines qui correspondent à ces deux valeurs, la colère, le ressentiment et la rage, ainsi que le remord et la culpabilité, impliquent donc une croyance en la responsabilité morale, comme nous venons de le proposer.

Dans le cas de la colère, Roberts donne l'exemple d'un couple qui a pour habitude de se répartir les tâches ménagères en divisant le travail au cours de la semaine. Karla rentre le mardi soir, jour de ménage de son conjoint, et découvre l'appartement en désordre. Agacée, elle s'empresse de ranger les lieux, alors que son mari, attablé, continue de siroter une bière. Au lieu de s'excuser et de proposer son aide, Günther fait preuve d'une dérision inadaptée et finit par se moquer d'elle franchement. Elle se met en colère, frappe du point sur le comptoir et lui reproche son comportement en hurlant.

Selon Roberts, toute colère, à l'image de cet exemple commun, implique une représentation (parfois erronée) de la responsabilité. La colère ne porte pas uniquement sur l'insatisfaction d'un désir mais représente, en plus, la personne sur laquelle l'émotion est portée, comme responsable de l'état de fait indésirable :

"The attribution of offense is of a special moral kind: He has not just hurt and inconvenienced her but has done so as a responsible agent. If she came to see the unfairness and mischief he has occasioned as arising from a pathology that completely bypassed his responsible agency (say, a brain tumor of just the right kind), then her emotion might be irritation, frustration, or disappointment, but it would not be anger."¹¹

¹¹ Roberts, R. C. (2003). *Emotions: An essay in aid of moral psychology*. Cambridge University Press

Par ailleurs, cette représentation transparaît aussi dans le type de désirs qui accompagnent spontanément la colère de Karla. Si Günther est un être responsable, il a délibérément mal agi et mérite donc d'être puni pour ses actes et omissions. Le comportement de Karla semble, à deux reprises, exprimer ce désir de punition qui présuppose la responsabilité de son mari. En effet, le lien naturel entre colère et désir consécutif de rétribution confirme l'attribution de responsabilité morale inhérente à l'émotion :

"Anger's consequent concern is also clearly illustrated in Karla's anger. When she finds Günther sitting at the table kibitzing about her industriousness, she first goes into the bedroom and starts cleaning up. She seems to intend to shame Günther by casting herself in a martyrial role. Then she changes strategy, taking the more direct approach of yelling and creating an embarrassing spectacle in front of their flatmates. Common to these two approaches is the intention to punish, to give deserved pain, to pay Günther back for his offenses. This is perhaps not quite generalizable because in some cases the angry person does not wish to do the punishing herself; so, we should say that anger's consequent concern is a desire that the offender be punished."¹²

Dans le cas du remord également, Roberts note, en premier lieu, le caractère moral de l'émotion qu'il tente de décrire. Il existe une différence fondamentale entre le regret et le remord et celle-ci se situe précisément au niveau de la responsabilité. Le regret est une attitude proche du remord, à la différence néanmoins significative que celui-ci n'implique pas la croyance d'avoir commis une faute. Nous regrettions souvent des états de fait sur lesquels nous ne reconnaissions avoir aucun contrôle. Les objets naturels du regret sont, de ce fait, très nombreux et divers. Je peux regretter la faim dans le monde, la souffrance d'un proche, ou encore le mauvais temps qu'il fait aujourd'hui. À l'inverse, le remord est une attitude qui nous lie intimement à l'état de fait sur lequel porte l'émotion. Dans le remord, la personne interprète son action comme une violation des normes qu'elle se fixe elle-même. Mais en plus d'en être la cause, la personne interprète cette faute comme étant le fruit d'un processus intentionnel et libre. Elle croit être responsable et estime mériter d'être blâmée pour son manquement :

"She construes the action or omission as not only her doing, but as something she did intentionally or negligently, and thus blameworthily."¹³

Ainsi, le remord implique une responsabilité morale qui ne figurait pas dans le regret. Toutefois, il arrive fréquemment que ces deux émotions soient confondues. Nous ressentons parfois du remord dans des situations où nous devrions ressentir du regret. Roberts évoque, à cet égard, le cas d'une automobiliste qui renverse un enfant sur la route, un jour de neige. Celle-ci conduisait à une vitesse adaptée aux conditions météorologiques et n'aurait rien pu faire pour éviter l'accident. Elle ressent

¹² Ibid.

¹³ Ibid.

néanmoins un remord irrationnel dans la mesure où elle est incapable de se concevoir autrement que responsable :

"Her remorse is irrational; she should be feeling intense regret instead. Perhaps she says to herself such things as "If only I had waited another five minutes to go out for milk," and so by a self-flagellating sophistry she makes it appear to herself that she is to blame. That she feels remorse, rather than regret, is determined by the terms in which she sees herself and the situation: in terms of her blameworthy fault."¹⁴

Si l'automobiliste ne ressent pas de regret, c'est précisément parce qu'elle est incapable de se concevoir autrement que comme un agent responsable. La colère, le remord, ainsi que toutes les attitudes réactives que nous avons passées en revue semblent donc impliquer nécessairement ce type de croyance en la responsabilité morale. C'est en ce sens précis que je caractérise les attitudes réactives comme des émotions morales¹⁵.

En résumé, dans cette partie, j'ai défini les attitudes réactives, en tâchant de montrer le lien fort qu'elles entretiennent avec l'attribution de responsabilité morale. Je vais, désormais, en dire un peu plus sur la responsabilité morale.

¹⁴ Ibid.

¹⁵ Fischer et Ravizza (1993) appellent ces attitudes réactives « responsibility-reactions »

1.2 Deux Approches de la Responsabilité Morale

Qu'entend-on exactement par responsabilité morale ? Selon la plupart des philosophes, la responsabilité morale désigne une relation entre un sujet et un évènement qui permet de justifier le blâme ou la louange.¹⁶ Comme les exemples de Karla et l'automobiliste le suggèrent, il semble que nous avons effectivement une propension naturelle à blâmer et louer certains agents. Plus que de simples croyances, ces jugements moraux sont reliés à des attitudes qui ne sont appropriées que si elles portent sur les agents adéquats. Il est donc plus approprié d'être en colère contre son supérieur hiérarchique, plutôt que son chat, en raison de la responsabilité morale qui fait défaut au second.

Cependant, que peut-on dire de la nature précise de ce lien entre attitudes réactives et responsabilité morale ? Est-ce la croyance en la responsabilité morale qui fonde ces émotions ou inversement, ces émotions qui fondent la croyance ? En des termes plus concrets, est-ce l'attribution de responsabilité qui fonde la colère ou inversement, la colère qui fonde la responsabilité ? Ces deux façons de concevoir le lien causal entre responsabilité morale et attitudes réactives nous importent car elles donnent lieu à deux définitions distinctes de la responsabilité morale. En effet, selon une première perspective, la responsabilité morale est définie indépendamment des réponses émotionnelles qu'elle suscite et se fonde sur des conditions rationnelles propres. Il s'agira donc de faire appel à des propriétés externes à nos réactions, pour concevoir correctement la responsabilité morale.

Selon la perspective alternative, la responsabilité morale est définie comme ce qui suscite une réponse émotionnelle, c'est-à-dire ce pour quoi les gens manifestent typiquement de la colère. Le concept ne requiert aucune propriété distincte des émotions elles-mêmes. En somme, être responsable n'implique qu'une disposition à recevoir les réactions émotionnelles exprimées par les autres.

À ce stade préliminaire, il n'est pas certain que l'une des deux explications bénéficie d'une priorité métaphysique sur l'autre : il est difficile de savoir en effet s'il faut plutôt expliquer l'attribution de responsabilité morale, à partir de la responsabilité morale ou alors l'inverse.

Néanmoins, je partage l'intuition de la thèse traditionnelle selon laquelle la responsabilité ne dépend pas de nos attributions, parce nous attribuons parfois de la responsabilité morale à des agents qui ne

¹⁶ Chevarie-Cossette, S.-P. (2017), *Responsabilité morale (A)*, dans Maxime Kristanek (dir.), l'Encyclopédie philosophique

sont pas véritablement responsables. Une analogie intéressante à l'appui de cette idée est celle de la plaisanterie. Il semble en effet que le fait qu'une plaisanterie soit drôle ou non ne dépende pas de la disposition manifestée par la majorité des gens à en être amusés. Il n'est effectivement pas impossible que le monde entier soit amusé par une blague qui n'est pas drôle. De la même manière, il est possible d'être tenu responsable sans l'être vraiment.

1.3 Thèse Traditionnelle de la Responsabilité Morale¹⁷

Selon les thèses traditionnelles de la responsabilité morale, les attitudes réactives ne sont que des corollaires émotionnels causés, en premier lieu, par nos attributions de responsabilité morale. La responsabilité morale est donc conçue comme fondamentalement distincte et antérieure aux attitudes réactives. Juger qu'une personne est responsable d'une offense à mon égard est une attitude distincte de ma colère envers cette même personne. Par conséquent, il est logiquement possible de juger qu'une personne ait responsablement commis un acte répréhensible, sans nécessairement ressentir d'indignation ou de désapprobation à son égard, bien que cette réaction soit simultanément causée, la plupart du temps. Selon cette théorie, c'est la responsabilité morale qui jouit d'une priorité métaphysique ou explicative sur les engagements émotionnels qui en découlent.

Mais sur quelle base cette première attribution de responsabilité morale est-elle, quant à elle, fondée ? Selon la plupart des philosophes, la responsabilité morale est un concept semblable aux autres, qui peut être analysé rationnellement en termes de conditions nécessaires et suffisantes.¹⁸ Il existe donc des conditions de la responsabilité morale qui peuvent être satisfaites ou non dans certains cas donnés. La méthode consacrée des philosophes s'étant penchés sur la question, au XXème et XXIème siècles, respecte souvent le même schéma : 1) identifier les conditions nécessaires et suffisantes au concept de responsabilité morale, 2) Appliquer ces conditions à des cas particuliers, 3) réviser possiblement les conditions établies en fonction des intuitions qui surgissent des cas étudiés.¹⁹ Bien que l'analyse des conditions de la responsabilité morale ait fait l'objet de nombreuses discordes philosophiques, il est néanmoins possible d'identifier des présuppositions communes aux différentes théories. En effet, la responsabilité morale nécessiterait, au moins, deux conditions évidentes qui figuraient déjà dans l'analyse aristotélicienne. Selon Aristote, un agent moral est toujours en mesure d'être blâmé ou loué pour ses actes, sauf s'il se trouve dans une situation de contrainte ou

¹⁷ Je tire ce terme de : Watson, G. (1987). *Responsibility and the Limits of Evil: Variations on a Strawsonian Theme*. Schoeman (Ed.), Responsibility, Character, and the Emotions. Cambridge University Press

¹⁸ Je pense, ici, à l'écrasante majorité des philosophes du Free will, compatibilistes et incompatibilistes, qui partent toujours d'une certaine analyse de la responsabilité morale pour se demander, dans un deuxième temps, si celle-ci est compatible avec le déterminisme.

¹⁹ Nous pouvons citer, comme exemple, la littérature très fournie des cas Frankfurt.

d'ignorance. La contrainte et l'ignorance sont ainsi identifiées par la théorie, comme des exceptions à la responsabilité morale :

« Puisque la vertu a rapport à la fois à des affections et à des actions, et que ces états peuvent être soit volontaires, et encourir l'éloge ou le blâme, soit involontaires, et provoquer l'indulgence et parfois même la pitié, il est sans doute indispensable, pour ceux qui font porter leur examen sur la vertu, de distinguer entre le volontaire et l'involontaire ; et cela est également utile au législateur pour établir des récompenses et des châtiments. On admet d'ordinaire qu'un acte est involontaire quand il est fait sous la contrainte, ou par ignorance. Est fait par contrainte tout ce qui a son principe hors de nous, c'est-à-dire un principe dans lequel on ne relève aucun concours de l'agent ou du patient : si, par exemple, on est emporté quelque part, soit par le vent, soit par des gens qui vous tiennent en leur pouvoir. (...) L'acte fait par ignorance est toujours non volontaire ; il n'est involontaire que si l'agent en éprouve affliction et repentir. »²⁰

En d'autres termes, la première condition d'Aristote est relative au contrôle. Pour qu'un agent soit moralement responsable de ses actes, il faut que celui-ci ne soit pas en proie à une force extérieure à lui-même. Il faut donc que l'agent bénéficie d'un degré de contrôle agentiel suffisant sur ses actions pour être jugé responsable. La deuxième condition est, quant à elle, de nature épistémique. Pour qu'un agent soit responsable de ses actes, il est également nécessaire que celui-ci sache ce qu'il est en train de faire et qu'il agisse en connaissance de cause.

Si ces deux conditions sont partagées de tous, c'est principalement parce qu'elles jouissent d'un soutien intuitif difficilement contestable. Pour ce qui est du contrôle, les phénomènes de menace, d'addiction, de coercition ou de manipulation sont des cas intuitivement limitants. En effet, une décision prise sous la contrainte d'une arme à feu paraît raisonnablement manquer de contrôle suffisant pour compter comme une action libre. Il peut exister, néanmoins, des situations plus difficiles à déterminer. Qu'en est-il d'une action qui est le fruit d'une certaine éducation ? Est-elle nécessairement une action contrainte ?

Concernant la condition épistémique, l'ignorance constitue également, selon Aristote, une limitation de la responsabilité morale. Imaginons un automobiliste qui tuerait un chaton caché dans l'une des roues de son véhicule, à l'occasion d'une marche arrière malheureuse.²¹ Il paraîtrait intuitivement inapproprié de le tenir responsable de cet acte et de considérer qu'il doit être blâmé.

En effet, une connaissance supplémentaire sur la localisation du chaton lui faisait indubitablement défaut au moment de la manœuvre.

²⁰ Aristote, *Ethique à Nicomaque*, 1109b30 – 1111b5

²¹ Exemple tiré de : Fischer, J. M., & Ravizza, M. (1993). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press

Son ignorance semble, par conséquent, l'excuser entièrement de son acte. Ainsi, pour être responsable d'une mauvaise action, il apparaît nécessaire de connaître les raisons qui en font une mauvaise action. C'est pour cette raison que le recours à l'ignorance est une excuse relativement commune dans nos conflits moraux. Les énoncés tels que « Je ne savais pas. » ou « Je ne pensais pas que ça te ferait du mal. » sont des tentatives d'excuses courantes qui semblent s'appuyer sur ce principe. Cette croyance repose certainement sur l'intuition fondamentale de l'incompatibilité entre ignorance et responsabilité.

Si la majorité des philosophes s'accordent aisément sur la pertinence des deux conditions que je viens de présenter, des dissensions théoriques apparaissent, néanmoins, dès lors qu'il s'agit d'en définir exactement les implications. En effet, bien que la condition épistémique soit relativement transparente, la condition de contrôle, quant à elle, semble requérir une explicitation supplémentaire. Quel est, pour ainsi dire, le type de contrôle nécessité par la responsabilité morale ? Quel degré de contrôle agentiel est-il suffisant pour mériter d'être blâmé ou loué ?

La réponse classique à cette question fait souvent appel au concept de libre arbitre. Il faut être libre pour être moralement responsable et mériter d'être blâmé ou loué. Cependant, si l'on se limite à cette première réponse, il apparaît que nous n'avons encore rien dit à propos du contrôle. Si la responsabilité morale dépend du libre arbitre, la question centrale n'est que déplacée. En effet, de quel type de contrôle parle-t-on en mobilisant ce nouveau concept ?

Il existe essentiellement, dans cette perspective, deux conceptions du libre arbitre, la première reposant sur **la possibilité alternative** (leeway) et la deuxième sur **la source agentive de l'action** (sourcehood).

Selon le principe de la possibilité alternative, pour être libre et moralement responsable d'une action A, dans une situation S, il faut être capable d'agir autrement dans S. Ainsi, un électeur américain est libre et moralement responsable de son vote pour Donald Trump, si et seulement s'il pouvait faire autrement dans l'isoloir, c'est-à-dire voter pour Joe Biden. Si ce même électeur se trouve sous la menace d'une arme à feu au moment de son choix, il apparaîtra alors que le principe de la possibilité alternative ne sera pas satisfait car, contraint, l'électeur ne pouvait raisonnablement pas voter autrement. Celui-ci n'est donc pas responsable de son vote. Une fois cette caractérisation de la responsabilité morale proposée, il convient toutefois de préciser davantage ce que l'on entend par possibilité d'agir autrement. Cette condition se limite-t-elle à l'absence de phénomènes de contraintes évidents, comme le pensait Aristote, ou est-elle à concevoir en termes strictement métaphysiques ? Autrement dit, est-il nécessaire au libre arbitre que le futur soit réellement ouvert à différentes possibilités contradictoires ?

Dans la deuxième conception du libre arbitre, c'est la capacité de générer une action (ou d'en être sa source) qui garantit un contrôle suffisant sur nos actions pour la responsabilité morale. Dans le vaste ensemble de ce que l'on pourrait appeler les *activités humaines*, il y aurait une différence pertinente entre celles dont nous reconnaissions être la source agentive (aller se promener, se servir une tasse de café, prendre une douche, etc...) et celles dont nous ne reconnaissions pas être entièrement la source agentive (grelotter lorsqu'il fait froid, faire battre son cœur, transpirer, etc...). Dans le premier cas, il nous semble produire l'action alors que dans le second, nous la subissons. Ainsi, selon cette distinction, le libre arbitre interviendrait uniquement dans les cas où nous sommes à *la source* de l'action, *être à la source* d'une action étant défini comme un pouvoir d'initiation volontaire de cette même action. Il est à noter que ces deux différentes caractérisations du libre arbitre, en termes de possibilité alternative et source agentive, ne sont pas nécessairement exclusives l'une de l'autre. Au contraire, il est même probable qu'elles s'impliquent mutuellement.

Cependant, la menace du déterminisme dans l'élaboration d'une définition adéquate du libre arbitre et de la responsabilité morale a été volontairement tue jusqu'ici. En effet, les deux conceptions du libre arbitre que je viens d'explorer sommairement pourraient nécessiter la falsification du déterminisme compris comme la thèse selon laquelle la conjonction des lois de la nature et l'ensemble des événements passés (c'est-à-dire de l'existence d'un état actuel du monde) n'implique qu'un seul et unique futur possible. Je reviens sur cette question dans le chapitre suivant (**2.1.**) et tâcherai de montrer précisément comment cette thèse met en cause les deux conditions du libre arbitre que je viens d'expliciter.

Chapitre II : Scepticisme à Propos des Thèses Traditionnelles de la Responsabilité Morale

Après avoir conçu les attitudes réactives comme des émotions morales et présenté la définition traditionnelle de la responsabilité en termes de contrôle, je m'intéresse, dans ce chapitre, aux conditions que nous venons d'identifier. Si le libre arbitre est défini comme un contrôle agentiel nécessaire à la responsabilité morale, reposant sur le principe des possibilités alternatives et/ou la source agentive, il convient encore de se demander si ces deux conditions sont remplies dans notre monde, sans quoi, les attitudes réactives ne sont plus justifiées.

En effet, si le type de contrôle agentiel requis par les attitudes réactives est impossible ou, en d'autres termes, si nous n'avons pas de libre arbitre, il apparaîtra, conséutivement, qu'elles sont toujours irrationnelles, ce qui pourrait représenter une conséquence problématique au niveau de nos vies affectives. Aussi, dans ce chapitre, je discuterai du problème du libre arbitre en me demandant si les conditions du libre arbitre peuvent être remplies. La menace classique qui nous permet d'en douter est celle du déterminisme. Mais le déterminisme invalide-t-il vraiment la responsabilité morale telle que nous venons de la concevoir ?

Dans un premier temps, je définirai cette thèse et montrerai en quoi elle peut incarner une menace directe pour la conception traditionnelle de la responsabilité morale (**2.1**). Ensuite, je présenterai les différentes positions classiquement défendues dans le débat sur la compatibilité entre libre arbitre et déterminisme (**2.2**). Puisque ma préoccupation première porte sur la responsabilité morale, je m'intéresserai principalement aux trois thèses qui garantissent l'existence du libre arbitre et me demanderai si leur position est cohérente.

Ainsi, je commencerai par présenter le compatibilisme (**2.3**) et soutiendrai que cette position est finalement incapable de résister à des arguments incompatibilistes décisifs (**2.4**).

Ensuite, je m'intéresserai aux deux formes de libertarianisme (**2.5**) et montrerai également que leur position n'est pas tenable à cause des incohérences qu'elles abritent dans les modèles (**2.6**).

Enfin, Je tirerai de ces réflexions la conclusion sceptique selon laquelle aucune entreprise théorique issue de la conception traditionnelle n'est à même de garantir véritablement la responsabilité morale (**2.7**).

2.1 Déterminisme et Incompatibilisme Primitif

Le déterminisme est la thèse selon laquelle tout évènement actuel (qu'il soit un élément physique ou mental²²) est la conséquence nécessaire des lois de la nature et des évènements antécédents²³.

En d'autres termes, la conjonction des lois de la nature et l'ensemble des évènements passés n'implique qu'un seul et unique futur possible²⁴. L'un des corollaires principaux de cette thèse est le suivant : si nous connaissons avec certitude les lois de la nature ainsi qu'un état particulier du monde, il serait théoriquement possible d'en déduire entièrement le futur et d'en reconstituer le passé.²⁵

Bien qu'il soit difficile d'expliciter ce que nous entendons précisément par « lois de la nature »²⁶, il est suffisant, dans le cadre de cette réflexion, de s'en tenir à une compréhension élémentaire selon laquelle il existe un ensemble contingent de propositions toujours vraies, partout dans notre univers.²⁷ Ces lois se conçoivent, ainsi, comme la somme des régularités objectives qui structurent notre monde. En guise d'exemples classiques, nous pouvons citer, par exemple, les lois de la gravitation qui expliquent la mécanique des corps lourds ou encore les lois physiques qui expliquent la transformation de l'énergie cinétique lors d'un impact.

Toutefois, croire au déterminisme n'implique pas que nous disposions d'une science intégrale de toutes les lois de la nature, puisque la vérité d'une chose ne dépend pas de sa description. Il est même probable que certaines lois soient si complexes que nous n'en connaissons que des approximations ou des descriptions partielles. Le déterminisme n'implique donc pas la totale connaissance des lois naturelles mais presuppose seulement l'existence d'un tel ensemble. Dès lors, si le monde est déterministe, l'ensemble des lois de la nature ne se limite pas seulement aux lois physiques que nous connaissons, mais comprend également les lois inconnues qui sont potentiellement à l'œuvre dans les mécaniques mentales complexes, comme la délibération humaine, par exemple. En effet, notre incapacité à identifier les causes d'un évènement donné n'implique pas l'inexistence de celles-ci.

²² À supposer que le second ne se réduise pas au premier.

²³ Hoefer, C. *Causal Determinism*. The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds)

²⁴ Van Inwagen, P. (1983). *Essai sur le libre arbitre*. (Trad. F. Nef). Vrin. (Traduction française de l'ouvrage "An Essay on Free Will")

²⁵ Cette idée fameuse, issue du déterminisme dur, doit son origine au « génie de Laplace », expérience de pensée formulée par le philosophe dans son *Essai philosophique sur les probabilités*.

²⁶ La définition de « loi de la nature » fait effectivement l'objet d'un large débat scientifique.

²⁷ On trouve cette caractérisation minimale chez Van Inwagen P. (1983)

Imaginons un grain de sable que nous prélevions aléatoirement parmi tous les grains de sable qui constituent le désert du Sahara. Bien qu'il paraisse évident que nous ne serons jamais en mesure de retracer toutes les causes qui déterminent sa position actuelle par rapport aux autres, nous ne postulons pas l'inexistence de ces causes pour autant. Au contraire, nous savons que la localisation du grain de sable est soumise à une nécessité causale qui remonte en dernier lieu aux conditions initiales de notre environnement terrestre ainsi qu'aux lois fondamentales de la matière. Nous ne serons cependant jamais capables de formuler une explication exhaustive (Pourquoi ce grain de sable est-il à cette place plutôt qu'une autre ?), parce que l'ensemble des relations causales pertinentes, considérées ensemble est trop complexe à appréhender.

Analogiquement, notre ignorance de la nature des lois qui agissent dans la délibération humaine (Pourquoi tel agent fait-il tel choix plutôt qu'un autre ?), n'implique pas l'inexistence de ces causes. De la même manière qu'il existe la loi de la gravitation, il existerait des lois complexes relatives aux processus délibératifs humains. L'action humaine pourrait effectivement respecter une causalité déterministe dont nous ne connaissons cependant ni les lois ni les conditions pertinentes. Ainsi, selon cette perspective, nos actes quotidiens ne sont, en définitive, pas plus libres que le mouvement d'un grain de sable, mais sont simplement soumis à des lois et des conditions différentes. Si le déterminisme est vrai, nous, humains, sommes donc soumis à la même nécessité universelle qui détermine le reste de l'univers.

Une autre façon de caractériser le déterminisme a recours à une formalisation modale :

Il n'existe pas de monde possible dans lequel les lois de la nature, ainsi que l'état actuel des choses sont strictement identiques à notre monde et dans lequel le futur est différent du nôtre²⁸.

Selon cette définition, tout monde possible qui présenterait les mêmes conditions initiales et les mêmes lois de la nature est un monde identique au nôtre et son futur l'est également. Si les conditions de notre univers et nos lois de la nature étaient recréées dans un monde possible, l'ensemble des événements produits dans notre monde le serait également dans ce monde identique.

En outre, dans quelle mesure cette thèse est-elle à prendre en considération dans une réflexion sur la responsabilité morale ?

Dans l'histoire de la philosophie, le déterminisme a été conçu, dès son origine, comme incompatible avec le libre arbitre et la responsabilité morale. Cette incompatibilité a d'abord donné lieu à la formulation d'un problème classique qui opposait deux hypothèses contradictoires portant sur les actions humaines. Selon la première hypothèse, nos actions sont libres parce qu'elles prennent naissance dans notre volonté. Selon l'hypothèse rivale, puisque le déterminisme est universellement

²⁸ Van Inwagen, P. (1983). *Essai sur le libre arbitre*. (Trad. F. Nef). Vrin. (Traduction française de l'ouvrage "An Essay on Free Will")

vrai, nos actes sont en réalité gouvernés par une nécessité causale qui ne laisse pas de place à la liberté. Cette opposition prit l'apparence d'un dilemme qu'il était nécessaire de résoudre en penchant pour l'une ou l'autre des deux explications exclusives :

A. Notre volonté est-elle libre ou est-elle déterminée ?

De nos jours, comme le souligne Peter Van Inwagen²⁹, cette façon d'envisager la dialectique entre libre arbitre et déterminisme semble avoir été remplacée par une formulation qui renvoie à une interrogation logiquement plus fondamentale que la première :

B. Libre arbitre et déterminisme sont-ils compatibles ?

La première formulation du problème du libre arbitre présupposait déjà, entre ces deux concepts, une incompatibilité qui nécessitait néanmoins une explication. En effet, on ne peut pas poser la question traditionnelle sans avoir déjà répondu, au moins implicitement, à la question de la compatibilité. Avant d'appréhender les différentes positions théoriques qui correspondent aux réponses possibles, je vais d'abord tâcher de mettre en lumière le type de réflexions qui a longuement tenu l'incompatibilité pour évidente. Si le déterminisme paraît fondamentalement en contradiction avec le libre arbitre, c'est principalement parce qu'il met en cause la vision selon laquelle le futur est ouvert à différentes possibilités alternatives. Or, puisque les deux conditions du libre arbitre (possibilité d'agir autrement et source agentive) prennent précisément racines dans une métaphysique permettant la réalisation de différents futurs possibles, le déterminisme, qui en est la négation, paraît intrinsèquement incompatible avec cette idée. Par conséquent, si ces deux conditions ne sont pas satisfaites, le libre arbitre ne peut pas exister dans un monde déterministe.

On désigne généralement cette position philosophique par « incompatibilisme ». Les tenants contemporains de cette thèse ont ensuite formalisé des arguments considérables à l'appui de l'intuition fondamentale que je viens de présenter. Considérons-en deux, dont j'ai simplifié la structure logique :

Argument contre la possibilité d'agir autrement :

1. *Personne n'a de contrôle sur les événements passés ni sur les lois de la nature.*
2. *Personne n'a de contrôle sur le fait que les événements passés ni les lois de la nature n'impliquent conjointement qu'un seul et unique futur possible.*

²⁹ Ibid.

3. *Donc, personne n'a de contrôle sur les évènements futurs.*³⁰

Selon cet argument, s'il ne peut y avoir qu'un unique futur possible déterminé par des facteurs qui nous dépassent, le principe des possibilités alternatives, qui requiert, quant à lui, une condition métaphysique selon laquelle le futur est ouvert à différentes possibilités que nous choisissons, est impossible. L'inévitabilité du futur invalide donc notre capacité à choisir réellement entre différentes options. Selon les incompatibilistes, la seule compréhension des exigences métaphysiques du principe des possibilités alternatives suffit à le rendre incompatible avec celles du déterminisme.

Argument incompatibiliste contre la source agentive :

1. *Une personne a un libre arbitre si et seulement si elle est la source ultime de son action.*
2. *Si le déterminisme est vrai, personne ne peut être la source ultime de son action, mais seulement la conjonction des évènements passés et des lois de la nature.*
3. *Donc, personne n'a de libre arbitre.*³¹

Selon ce second argument, localiser la source ultime d'une action au niveau de l'agent qui la produit n'aurait pas véritablement de sens, dans un monde déterministe. Si l'on prend en considération les implications strictes de cette thèse, il est même faux de prétendre que, dans l'historique causale qui a mené à une action, la source ultime, c'est-à-dire la première cause, se trouve dans l'agent. Ultimement, la source de l'action remonte aux lois naturelles et à un état de choses qui préexistait à notre naissance. La condition métaphysique pour que l'agent constitue la source ultime de son action exigerait que la chaîne causale s'arrête à la prise de décision de cet agent. Or, ceci ne serait possible que si l'agent avait la possibilité d'initier une chaîne causale à partir de rien, ce qui est, en principe, impossible dans un cadre déterministe. Là encore, le libre arbitre semble incompatible avec le déterminisme.

³⁰ Cet argument respecte de près la formulation de l'argument de la Conséquence proposé par Van Inwagen (1983).

³¹ Vihvelin, K. *Arguments for Incompatibilism*. The Stanford Encyclopedia of Philosophy (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.)

2.2 Le Problème du Libre Arbitre

Bien que l'incompatibilisme repose, comme je viens de l'esquisser, sur un soutien argumentatif relativement efficace de prime abord, il convient cependant de reposer la question de la compatibilité, afin d'obtenir un panorama général des thèses défendues. En effet, différentes réponses demeurent possibles. Voici la structure dialectique du problème du libre arbitre :

1. Le Compatibilisme est la thèse selon laquelle le libre arbitre et le déterminisme sont compatibles.

2. L'Incompatibilisme est la thèse selon laquelle le libre arbitre et le déterminisme sont incompatibles.

Il existe, dans l'incompatibilisme, deux positions opposées :

2.1. Le Libertarianisme est la thèse selon laquelle nous possédons un libre arbitre et le déterminisme est faux.

Cette thèse connaît deux formes distinctes :

2.1.1. Le libertarianisme de la causalité agentive³² est la thèse selon laquelle l'action libre n'est pas causée par les événements précédents, mais directement par l'agent.

2.1.2. Le libertarianisme de la causalité évènementielle³³ est la thèse selon laquelle l'action libre est causée par les événements précédents, mais à la suite d'un processus indéterministe.

2.2 Le déterminisme dur³⁴ est la thèse selon laquelle le déterminisme est vrai et nous ne possédons pas de libre arbitre.

Ainsi, de toutes les thèses défendues dans le débat entre libre arbitre et déterminisme, trois d'entre elles prétendent que le libre arbitre est possible (seul le déterminisme dur soutient le contraire).

Le compatibilisme, tout d'abord, prétend que le concept de libre arbitre est compatible avec le déterminisme : libre arbitre et déterminisme peuvent donc être vraies simultanément.

Les deux formes de libertarianisme sont des thèses incompatibilistes qui prétendent que, puisque le déterminisme n'est pas vrai, le libre arbitre et la responsabilité morale sont saufs. Elles partagent donc avec le compatibilisme une croyance commune en l'existence de la responsabilité morale.

Mais comment ces trois thèses sont-elles en mesure de garantir une telle croyance ?

³² De l'anglais « agent-causal libertarianism »

³³ De l'anglais « event-causal libertarianism »

³⁴ De l'anglais « Hard determinism », terme emprunté à Pereboom (2003).

2.3 Compatibilisme

L'objectif de toute théorie compatibiliste est de proposer une analyse du libre arbitre, en termes de possibilités d'agir autrement et/ou de source agentive, qui soit en mesure de résister aux arguments incompatibilistes que nous avons vus plus haut. Mais comment peut-on défendre à la fois que le déterminisme est vrai et que nous sommes libres ? N'y a-t-il pas une contradiction dans les termes ? La meilleure façon de contrer cette intuition est de montrer, au moyen d'analyses conceptuelles alternatives, que le libre arbitre ne nécessite pas de conditions métaphysiques aussi strictes, à savoir que le futur soit ouvert à différentes possibilités et que l'agent soit capable d'initier une chaîne causale.

En effet, il semble que la réponse à la question de la compatibilité entre déterminisme et libre arbitre dépende essentiellement de ce que l'on entend par « être libre », ou plus exactement, par « possibilité d'agir autrement » et « source agentive de l'action », car ces notions ne sont peut-être pas si transparentes qu'elles le paraissent. Ainsi, si les compatibilistes parviennent à montrer que leur analyse du libre arbitre ne nécessite pas de conditions métaphysiquement trop coûteuses, ils auront alors en leur possession un soutien fort pour établir la compatibilité entre ces deux concepts.

Selon les compatibilistes classiques, s'inscrivant dans la lignée de David Hume, la liberté peut être analysée de façon conditionnelle plutôt qu'en termes de possibilités robustes. La liberté est avant tout une notion ordinaire qui désigne la capacité à « faire ce que l'on veut ». Or, même si le déterminisme était vrai, il subsisterait néanmoins un sens selon lequel nous continuons à avoir des désirs et à faire « ce que nous voulons ».

En effet, « Je lève le bras, *si j'ai le désir de lever le bras.* » semble intuitivement une bonne description d'action libre. Bien qu'il soit déterminé que je veuille lever le bras et que je le lève en conséquence, cette nécessité causale ne porte aucun préjudice au lien de dépendance conditionnelle entre ma volonté et mon action. Il reste correct, en effet, de dire « *si je n'avais pas formé le désir de le lever, je n'aurai pas levé mon bras.* », et ce même s'il était déterminé que je le veuille.³⁵

Dans cette approche, la possibilité d'agir autrement est analysée conditionnellement à nos désirs. Le sens profond du libre arbitre ne tiendrait finalement pas tant à la possibilité robuste d'agir autrement, qu'au lien fort entre nos volontés et nos actions. Ainsi, par la conservation de ce lien conditionnel, il subsiste un sens compatibiliste selon lequel, pour toute situation donnée, nous aurions pu faire autrement, *si nous avions voulu faire autrement.*

³⁵ Guillon, J.-B. (2016). *Libre Arbitre (GP)*. Dans Maxime Kristanek (dir.), l'Encyclopédie philosophique.

Il est à noter, par ailleurs, que cette conception du libre arbitre permet des distinctions intuitives entre les actions dérivant de notre volonté, et celles qui sont contraintes. Un prisonnier par exemple, ne disposerait pas d'un libre arbitre conditionnel, dans la mesure où *vouloir faire autrement* ne lui suffit pas pour *faire autrement* dans sa situation. La raison en est que le lien conditionnel entre volonté et action est politiquement supprimé dans son cas.

Néanmoins, cette analyse conditionnelle de la possibilité d'agir autrement se heurte à une difficulté majeure. En effet, il semble parfois qu'elle permette des conséquences inacceptables, en ce qu'il est possible d'attribuer la possibilité d'agir autrement à une personne qui, au moment de son action, n'avait pertinemment pas d'autre alternative possible. Ce point a été mis en valeur dans l'expérience de pensée suivante :

"Suppose that Danielle is psychologically incapable of wanting to touch a blond-haired dog. Imagine that, on her sixteenth birthday, unaware of her condition, her father brings her two puppies to choose between, one being a blond-haired Lab, the other a black-haired Lab. He tells Danielle just to pick up whichever of the two she pleases and that he will return the other puppy to the pet store. Danielle happily, and unencumbered, does what she wants and picks up the black Lab. When Danielle picked up the black Lab, was she able to pick up the blond Lab? It seems not. Picking up the blond Lab was an alternative that was not available to her. In this respect, *she could not have done otherwise*. Given her psychological condition, she cannot even form a want to touch a blond Lab, hence she could not pick one up. But notice that, *if* she wanted to pick up the blond Lab, *then she would have done so*. Of course, if she wanted to pick up the blond Lab, then she would not suffer from the very psychological disorder that causes her to be unable to pick-up blond-haired doggies. The classical compatibilist analysis of 'could have done otherwise' thus fails. According to the analysis, when Danielle picked up the black Lab, she *was able* to pick up the blonde Lab, even though, due to her psychological condition, she *was not able* to do so in the relevant respect. Hence, the analysis yields the wrong result."³⁶

Cette incohérence de l'analyse conditionnelle a eu pour conséquence radicale le rejet en bloc du principe de la possibilité alternative par la plupart des compatibilistes contemporains qui lui ont finalement préféré de nouvelles conditions du libre arbitre. La transition a conjointement été rendue possible par les travaux d'Harry Frankfurt qui ont jeté un véritable discrédit sur la pertinence même du principe dans le débat.³⁷ Frankfurt proposa alors une nouvelle analyse du libre arbitre qui place la question de la source agentive au centre. En effet, il peut exister une interprétation compatibiliste de la source agentive qui ne nécessite pas un pouvoir agentiel ultime, contrairement à ce que prétendent les incompatibilistes. Selon cette interprétation, ce qui importe avant tout, c'est que la personne libre

³⁶ McKenna, M. et Coates, D. J. *Compatibilism*. The Stanford Encyclopedia of Philosophy (Fall 2021 Edition), Edward N. Zalta (ed.)

³⁷ J'éviterai, ici, de me plonger dans les discussions très complexes autour de la cohérence des cas Frankfurt et tiendrais pour acquis que ceux-ci permettent réellement de douter du bienfondé du principe, en montrant que, dans certaines situations possibles, un agent est responsable de son action même s'il n'avait pas la possibilité d'agir autrement.

ait la capacité de s'identifier à ses propres choix et de les considérer comme les siens. "Être à la source" d'une action désignerait alors une compatibilité entre les traits de caractères d'une personne et les actions qu'elle produit, plutôt qu'une stricte capacité d'initiation causale. On nomme cette approche « thèse de l'identification ». D'après celle-ci, la source agentive n'est finalement pas une condition incompatible avec le déterminisme :

"On compatibilist identification accounts, what matters for self-determination is not whether our actions are determined or undetermined, but whether they are brought about by motives with which the agent is identified: exercises of the power of self-determination consists in an agent's actions being brought about, in part, by an agent's motives with which she is identified."³⁸

Selon les compatibilistes classiques, pour qu'un agent s'identifie à son action, il suffit que celle-ci soit la conséquence de sa volonté la plus forte. La plupart des compatibilistes contemporains seraient néanmoins en désaccord avec cette analyse simpliste. En effet, cette thèse ne permet pas de rendre compte des comportements compulsifs exprimés par certaines personnes. Dire qu'un agent est à la source de son action, sitôt que sa volonté la plus forte est déterminante, aurait pour conséquence pratique, par exemple, qu'une personne victime d'une sévère addiction consommerait librement des substances, simplement parce que son action est le résultat de sa plus forte volonté.

Or, ces conséquences théoriques paraissent difficilement acceptables parce qu'elles presupposent d'emblée l'impossibilité de contraintes psychologiques sur nos désirs. De plus, il semble parfois que notre volonté la plus forte ne soit pas nécessairement celle qui nous représente le mieux.

Toutefois, il est encore possible d'affûter la thèse de l'identification pour qu'elle soit en mesure de répondre à ce problème, tout en sauvegardant l'intuition première :

"Many compatibilists have developed identification accounts of self-determination that attempt to draw a distinction between an agent's desires or motives that are internal to the agent and those that are external.

The idea is that while agents are not (or at least may not be) identical to any motivations (or bundle of motivations), they are *identified* with a subset of their motivations, rendering these motivations internal to the agent in such a way that any actions brought about by these motivations are *self-determined*."³⁹

Ainsi, les volontés internes se distinguerait des volontés externes, d'une part parce que les secondes sont différentes de notre nature profonde, et, d'autre part, parce que nous nous identifions uniquement au premier type de volontés. « Être libre » se concevrait donc comme la capacité à agir conformément aux volontés auxquelles nous nous identifions seulement. Selon cette conception, une personne agoraphobe qui déciderait de ne pas se présenter à un anniversaire ne serait donc pas jugée

³⁸ McKenna, M. et Coates, D. J. *Compatibilism*. The Stanford Encyclopedia of Philosophy (Fall 2021 Edition), Edward N. Zalta (ed.)

³⁹ O'Connor T. et Franklin C. *Free Will*. The Stanford Encyclopedia of Philosophy (Winter 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.).

responsable de son omission. En effet, ne pas sortir de chez elle est peut-être sa volonté la plus forte, mais elle ne s'identifie pas nécessairement à cette volonté qui naît d'une contrainte psychique plus large.

Enfin, les thèses compatibilistes de l'identification sont intéressantes en ce qu'elles constituent une réponse à l'argument de la Conséquence. Selon celles-ci, nous sommes tenus responsables de nos actions, non pas en raison du contrôle que nous avons sur les facteurs causaux qui déterminent nos actions, mais plutôt en raison du contrôle qui naît de l'identification, à savoir de la correspondance entre une personne et ses actes. Sur cette base, il est possible de réfuter la conclusion de l'argument de la Conséquence selon laquelle aucune sorte de contrôle n'est possible si le déterminisme est vrai. Mais que cela signifie-t-il précisément pour un agent de s'identifier à ses actions et est-ce un contrôle suffisant pour être responsable ?

Je vais désormais présenter deux exemples influents de thèses compatibilistes qui poursuivent ces réflexions de manière ambitieuse. Ces deux candidats théoriques sont largement discutés en philosophie contemporaine.

2.3.1 La Thèse du Désir de Haut Niveau : Frankfurt

Selon la thèse d'Harry Frankfurt, l'identification repose essentiellement sur la distinction de deux types fondamentaux de désirs humains : les désirs de premier ordre et les désirs de second ordre. Un désir de premier ordre se forme à propos d'une action et se conçoit comme une motivation orientée vers la réalisation de cette action (*désirer écouter de la musique, désirer voyager, désirer jouer aux échecs*).

Un désir de second ordre, quant à lui, se forme à propos d'un désir de premier ordre et se conçoit comme un *désir de désirer*, qui remplit également une fonction motivationnelle (*désirer avoir le désir de lire Aristote, désirer avoir le désir de faire un marathon*).

Cette distinction structurelle constitue, selon Frankfurt, une particularité proprement humaine. En effet, les êtres humains formeraient deux types de désirs, contrairement aux animaux qui sont réduits au premier ordre. Cette distinction n'est toutefois pas toujours appréciée convenablement car, dans de nombreux cas, il y a conformité entre les deux niveaux : 1) Je désire boire un café, 2) je désire avoir le désir de boire un café et 3) Je bois un café sur la base de ces deux désirs.

Pour mieux saisir le sens de cette distinction, il est nécessaire de considérer les situations de conflits internes où les deux désirs ne coïncident pas nécessairement dans la psychologie d'un agent : un fumeur peut désirer fumer une cigarette, tout en ne désirant pas avoir ce désir. À l'inverse, un

étudiant de philosophie peut ne pas désirer lire l'œuvre complète de Kant, tout en désirant avoir ce désir.

Selon Harry Frankfurt, le libre arbitre intervient, dès lors que la fonction motivationnelle est exclusivement occupée par nos désirs de second ordre, plutôt que nos désirs de premier ordre. Si nos désirs de second ordre constituent essentiellement notre liberté, c'est parce qu'ils jouissent selon lui d'une supériorité hiérarchique et sont souvent le reflet de considérations profondes sur la direction de nos existences. Le libre arbitre ne consiste donc pas seulement en une simple capacité à avoir des désirs, mais surtout en notre capacité à nourrir des désirs profonds et auxquels nous accordons plus d'importance. En ce sens, la théorie de Frankfurt se rattache à la notion d'identification en ce que les désirs de haut niveau sont conçus comme les désirs auxquels nous nous identifions le plus. Ainsi, cette conception du libre arbitre a pour premier avantage de demeurer neutre quant à l'existence du déterminisme, puisque la liberté y est conçue comme le pur résultat d'une fonction humaine, évacuant ainsi la question métaphysique. En outre, la théorie a pour avantage de rendre possible une distinction importante entre, d'une part les actions nées d'une perte de contrôle et, d'autre part, les actions libres au sens explicité à l'instant.

En effet, un agent victime d'agoraphobie, par exemple, ne répond pas aux conditions de responsabilité morale proposées par Frankfurt. Lorsqu'un tel agent A est absent à un anniversaire, il possède 1) un désir de rester chez lui, mais 2) ne désire pas avoir le désir de rester chez lui et 3) son absence de désir de second ordre ne suffit pas à le motiver à se présenter à l'anniversaire. Il n'est donc pas libre de son action parce qu'elle échappe au contrôle normalement garanti par le désir de plus haut niveau.

Avec cette théorie, il semble que Frankfurt parvienne donc à proposer une analyse de la responsabilité morale qui satisfait deux desiderata compatibilistes principaux. Premièrement, la responsabilité morale étant conçue indépendamment de la question métaphysique, elle est, par définition, immunisée au déterminisme. Deuxièmement, la thèse de l'identification de Frankfurt nous permet néanmoins de faire des distinctions intuitives entre actions libres et contraintes.

2.3.2 La Thèse de la Sensibilité aux Raisons : Fischer et Ravizza

Selon cette seconde thèse de l'identification, le libre arbitre repose sur un mécanisme de raisonnement pratique que Fischer et Ravizza nomment la « sensibilité aux raisons »⁴⁰. En effet, lorsque nous délibérons, il apparaît que nos décisions soient motivées par nos différentes raisons d'agir évaluées lors du processus. Dans la mesure où notre capacité à réguler nos actions est garantie

⁴⁰ "Reason-responsiveness" tiré de : Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.

sur une base rationnelle, nous avons un libre arbitre parce que nous nous *identifions* à nos actions par le biais de nos raisons. En d'autres termes, par la délibération, nous faisons de nos actions les nôtres en considérant des raisons qui reflètent nos buts, motivations et valeurs qui sont intimement liés à la personne que nous sommes et à la personne que nous voulons être. Cette composante rationnelle du modèle permet la liberté, parce que l'agent agit conformément à ses raisons intimes, plutôt qu'en réagissant aveuglément à des facteurs externes.

Il existe deux formulations du modèle de la sensibilité aux raisons. La première est celle que Fischer et Ravizza appellent la version forte (« *strongly reason-responsive*⁴¹ ») :

*Un agent est fortement sensible aux raisons lorsqu'un certain type K de mécanisme, qui implique l'examen rationnel par l'agent des raisons pertinentes à la situation, donne lieu à une action, et s'il y avait des raisons suffisantes pour qu'il agisse autrement qu'il ne le fait réellement, il serait réceptif à ces raisons et aurait choisi et agi autrement par le pouvoir du même mécanisme délibératif qui aboutit à l'action.*⁴²

Cependant, cette version forte peut donner lieu à des conclusions contre-intuitives. En effet, ce modèle a pour conséquence qu'un agent qui ignorerait volontairement les raisons suffisantes pour agir autrement, ne serait pas moralement responsable. Or, ne pas vouloir accorder d'importance à de bonnes raisons d'agir autrement ne peut être un critère d'exclusion valable à nos obligations morales. Imaginons, par exemple, un menteur qui ignore les raisons qu'il a de dire la vérité. Au sens fort de la sensibilité aux raisons, le menteur n'est pas responsable de son acte puisque ses raisons de dire la vérité, bien que suffisantes, sont ignorées et n'ont pas de pouvoir motivationnel. En ce sens, il n'est donc pas sensible aux raisons. Or, il semble néanmoins que ce type de négligence ne puissent pas constituer une excuse valable à nos mauvais comportements. Fischer et Ravizza proposent alors une version faible :

*Un agent est faiblement sensible aux raisons lorsqu'un certain type K de mécanisme, qui implique l'examen rationnel par l'agent des raisons pertinentes à la situation, donne lieu à une action, et, au moins dans certaines circonstances alternatives dans lesquelles il existe des raisons suffisantes pour qu'il agisse autrement qu'il ne le fait réellement, il serait réceptif à ces raisons et aurait choisi et agi autrement par le pouvoir du même mécanisme délibératif qui aboutit à l'action.*⁴³

Selon ce second modèle, un agent A agit librement lorsqu'il vote pour Donald Trump, parce qu'il est sensible aux raisons, dans la mesure où, s'il avait eu à disposition de bonnes raisons de voter

⁴¹ Ibid.

⁴² Ibid.

⁴³ Ibid.

autrement, ces raisons lui auraient permis, dans certaines circonstances, d'agir autrement et de voter pour Joe Biden. Il est à noter que cette théorie permet, comme celle de Frankfurt, de maintenir une intuition forte quant aux phénomènes de contraintes. Elle maintient en effet l'intuition qu'un agoraphobe n'est pas libre, dans le sens explicité, parce que sa maladie réduit la fonction motivationnelle de ses raisons. Un agoraphobe n'est pas responsable de son absence à l'anniversaire, car il n'agit pas sensiblement aux raisons, dans la mesure où aucune raison suffisante ne lui aurait permis, dans aucune circonstance, de se présenter. Ses actions sont conditionnées par un mécanisme irrationnel qui rend donc l'identification impossible simultanément. En somme, Fischer et Ravizza, par le biais de cette thèse, proposent une analyse de la responsabilité morale compatible avec le déterminisme puisque la détermination causale ne nous empêche pas d'agir raisonnablement, bien au contraire.

2.4 Arguments contre le Compatibilisme

Comme nous venons de l'observer, le compatibilisme a su élaborer des analyses fines du libre arbitre qui résistent aux arguments incompatibilistes classiques. Pour ce faire, les compatibilistes réfutent, tout d'abord, le principe des possibilités alternatives, puis avancent, dans un second temps, une interprétation identificationiste de la source agentive. Toutefois, je tâcherai désormais d'argumenter que cette position peut être critiquée de façon décisive.

Tout d'abord, j'essaierai de montrer que le déterminisme implique la présence d'un hasard primitif qui semble fondamentalement incompatible avec la responsabilité morale. Cette première critique cherchera à motiver l'intuition selon laquelle la chance constitutive, en mettant la lumière sur l'origine hasardeuse de notre nature, remet en cause la pertinence des thèses de l'identification.

Dans un second temps, je défendrai que la définition compatibiliste du libre arbitre entre parfois en collision avec nos intuitions. En effet, si dans certaines situations de manipulation, une personne remplit toutes les conditions identificationiste du libre arbitre et n'est pas libre pour autant, il existera alors des contre-exemples solides pour affirmer que le type de contrôle évoqué par les compatibilistes n'est pas suffisant pour la responsabilité morale.

2.4.1 Chance Constitutive

Il existe, concernant le lien entre la chance et la responsabilité morale, un curieux paradoxe appelé « Chance morale »⁴⁴. Celui-ci naît d'une inconsistance qui intervient lors de nos jugements moraux. Tout d'abord, il semble que nous jugeons toute occurrence de chance dans la production de nos actions comme fondamentalement incompatible avec la responsabilité morale. Lorsque nous tenons une personne pour responsable de son acte, il semble en effet que nous presupposions d'emblée que l'acte a été produit d'une telle façon que ses intentions claires éliminaient la possibilité d'un hasard. C'est d'ailleurs cette intuition qui soutient la perspective kantienne selon laquelle les jugements moraux doivent porter essentiellement sur les intentions qui motivent l'action plutôt que l'action elle-même, celles-ci relevant du contrôle de l'agent, contrairement aux conséquences qui sont souvent hasardeuses et variables.

On peut facilement rendre compte de l'approche kantienne à l'aide de cas plus ou moins ordinaires.

⁴⁴ En anglais “Moral Luck” tiré de : Nagel, T. (1976). Moral luck. Proceedings of the Aristotelian Society, 50, 137-151. & Williams, B. (1981). *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press.

Imaginons, en effet, que Léa, soit un agent ordinaire, se promène dans la rue avec une pièce de monnaie dans la main. Elle décide de la lancer en l'air pour la faire tourner et se distraire. Par hasard, la pièce atterrit dans le chapeau d'un sans-abri assis sur le trottoir. Celui-là, surpris, regarde autour de lui et aperçoit Léa. Il la remercie beaucoup pour cette bonne action et l'informe qu'il utilisera cet argent pour acheter de la nourriture. Dans cette situation, bien que l'action ait entraîné une conséquence indubitablement positive, il est évident que Léa n'avait pas l'intention spécifique de faire un don au sans-abri. Il semblerait donc étrange de juger qu'elle mérite louange pour son acte hasardeusement altruiste. En effet, il semble clair qu'une action ne doit jamais être évaluée moralement en fonction du degré de chance ou de malchance qu'elle comporte. Lorsque nous louons une personne pour une bonne action, nous désirons au contraire que notre évaluation morale puisse rejoaillir d'une certaine manière sur le caractère à partir duquel l'action a été produite. Or, si l'action est hasardeuse, elle ne dit rien du caractère de l'agent.

Imaginons, à présent, que Léa souhaite aider un ami en prêtant sa voiture, pour qu'il puisse aller chercher ses enfants à l'école. Léa possède l'intention altruiste d'aider son ami en lui rendant un service. Cependant, en conduisant la voiture, l'ami se blesse gravement lors d'un accident provoqué par une chute de pierres hasardeuse. Bien que l'accident représente une conséquence regrettable, il semblerait encore une fois injuste de blâmer Léa pour avoir occasionné, par son acte, les blessures de son ami. À nouveau, la conséquence de l'action est ici le résultat d'un hasard qui échappe au contrôle de l'agent et qui, de ce fait, ne peut rejoailler correctement sur son caractère ou ses intentions. Après l'examen de ces deux exemples, il semble, à première vue, que la thèse kantienne décrive de manière satisfaisante nos mécanismes moraux intuitifs, car nous avons, en effet, tendance à accorder une valeur morale essentiellement déterminante aux intentions. Cette tendance conduit également à l'idée que nous devons, en jugeant exclusivement les intentions, protéger nos réactions morales de la mauvaise influence de toute forme de chance circonstancielle.

Toutefois, il apparaît, en réalité, que nos réactions morales ne sont, face à la chance, pas aussi insensibles qu'on pourrait l'espérer.

En effet, nous attribuons parfois au hasard plus de signification morale que nous voulons bien lui en reconnaître. Un des exemples classiques de chance morale est celui du « Lorry Driver ».⁴⁵

Considérons deux conducteurs de poids lourd qui, en état d'ivresse, empruntent la même route, l'un suivant l'autre. Le premier conducteur percute et tue accidentellement un enfant qui apparaît soudainement sous ses yeux. Dans ce cas, bien que les deux conducteurs aient eu la même intention répréhensible, seul le premier est véritablement perçu comme responsable de la mort de l'enfant.

⁴⁵ Williams, B. (1981). *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press.

Les deux conducteurs semblent donc soumis à un traitement moral différent, compte tenu des conséquences différentes qui dérivent néanmoins d'une intention partagée. Pourtant, nous reconnaissons volontiers que l'apparition soudaine de l'enfant sur la route, à ce moment précis, est une donnée hasardeuse qui échappait radicalement au contrôle des deux conducteurs. Si l'enfant était apparu au moment du passage du second camion, le second conducteur l'aurait percuté et serait donc responsable de sa mort. Ainsi, la conséquence hasardeuse de l'action bénéficie étrangement d'une valeur moralement déterminante.

Considérons aussi le cas d'un footballeur qui tire un penalty pour gagner un match crucial, mais, en le ratant, entraîne la défaite de son équipe. Si le joueur avait marqué le penalty, il serait considéré comme un héros. Cependant, en raison de sa malchance (Pour certaines raisons qui échappaient à son contrôle, le penalty a été manqué alors que le joueur en réussit une très grande majorité à l'entraînement), le joueur est considéré comme mauvais et suscite critiques et blâmes pour avoir occasionné la défaite de son équipe. Bien que le joueur ait eu l'intention de faire gagner son équipe en tirant le penalty, il est toutefois tenu responsable pour les mauvaises conséquences de son acte. On classe généralement ces deux exemples dans la catégorie de la « chance morale » qui décrit une situation paradoxale dans laquelle nous acceptons d'octroyer à la chance une signification morale que nous lui refusons en même temps. Pour dissoudre ce paradoxe, nous pouvons, en suivant la théorie kantienne, défendre qu'il s'agisse de deux exemples d'attribution morales erronées qui portent de façon illicite sur les conséquences hasardeuses de l'action, alors qu'elles devraient juger les intentions seulement. Résoudre le problème de la chance morale reviendrait alors à résoudre les erreurs d'attribution en appliquant strictement le modèle kantien. Mais pour que cette solution soit efficace, il faut encore expliquer en quoi nos intentions sont libres et donc susceptibles d'être blâmées et louées, contrairement à nos actions. Or, si le déterminisme est vrai, nous sommes précisément contraints de douter de ce présupposé selon lequel nos intentions sont produites librement. En effet, si nos intentions, au même titre que toutes les choses dans l'univers, sont déterminées par un ensemble de causes qui dépendent en dernier lieu des lois naturelles et du passé, il semble finalement qu'elles échappent à notre contrôle, de la même manière que la météo, le mouvement des astres et nos actions. Bien que nous agissions en suivant nos intentions propres, nous n'avons cependant aucun contrôle sur les causes de ces intentions elles-mêmes. Nos intentions sont le résultat de ce que Thomas Nagel nomme la « chance constitutive » d'un agent, c'est-à-dire le hasard qui est à l'origine des caractéristiques fondamentales d'un agent. Plus précisément, la chance constitutive détermine les traits de caractère, inclinations, tempéraments et dispositions avec lesquels nous naissions. Or, si la chance est à l'origine des traits innés qui constituent la personne que nous sommes, elle conditionne aussi les intentions que nous formons en conséquence. Imaginons un agent qui, par malchance,

possède une nature colérique. Ce tempérament inné détermine ensuite une plus grande fréquence d'intentions immorales qui, impliquant, par exemple, des comportements violents réguliers. Selon la thèse kantienne, il faut évaluer moralement cet agent fictif sur la base des intentions immorales qu'il formule. Mais, si nous punissons ce sujet colérique, nous acceptons simultanément que la punition condamne des intentions dont les origines sont hasardeuses et échappent à son contrôle. Nous acceptons donc de donner une signification morale à la chance et le paradoxe se trouve généralisé. Dès lors, si le déterminisme est vrai, nous devons considérer la chance constitutive et reconnaître qu'elle échappe à notre contrôle. Les deux thèses compatibilistes de Frankfurt et Fischer & Ravizza abritent, au même titre que la thèse kantienne, un paradoxe équivalent, dans la mesure où les désirs de haut niveau et les raisons d'agir d'un agent sont aussi soumis à la chance constitutive.

Bien que le lien d'identification entre nos actions et notre nature profonde soit garanti, ces thèses ne neutralisent pas l'influence de la chance au niveau de cette même nature, ce qui distille du hasard dans l'identification. Dans les thèses de l'identification, ce problème paraît néanmoins négligé, par définition, parce que le libre arbitre, tel qu'il est défini, n'a pas pour vocation de rendre la chance impossible. Selon les définitions observées, nous sommes libres de nos actions, même si le déterminisme est vrai et si la chance constitutive nous détermine, car nos actions trouvent toujours un ancrage en nous. Il est néanmoins possible de douter de la légitimité de cette approche en se demandant si les jugements moraux occasionnés sont justes. Selon la thèse de l'identification, nous pouvons blâmer et donc punir les agents qui s'identifient assez à leurs mauvaises actions. Or, si l'on admet en même temps que l'identification est fonction d'une nature hasardeuse et incontrôlée, il semble, en dernier lieu, que nous acceptons de punir les agents qui pâtissent du hasard et de récompenser ceux qui en bénéficiaient. Si le déterminisme est vrai, blâmer un agent pour la nature de ses désirs de haut niveau pourrait s'avérer aussi injuste que de le blâmer pour les proportions de son nez, et cette conclusion n'est raisonnablement pas souhaitable.

2.4.2 Argument de la Manipulation

Comme je l'ai présenté, l'identification est la voie privilégiée dans les analyses compatibilistes contemporaines de la responsabilité morale. Selon les thèses paradigmatisées de Frankfurt et Fischer & Ravizza, ce type de contrôle est compatible avec la vérité du déterminisme. Je vais désormais tâcher de montrer comment l'un des arguments incompatibilistes les plus efficaces parvient à rendre cette analyse insuffisante pour la responsabilité morale. On doit un fameux « argument de la manipulation » au philosophe Derk Pereboom qui, à l'occasion de trois expériences de pensée successives, met en scène des cas de manipulation qui vont du plus flagrant au plus atténué. Dans ces trois situations, la

manipulation implique la présence d'un manipulateur et d'une victime. Le manipulateur contrôle sa victime de la même manière qu'un ingénieur contrôle son robot, en s'assurant que celui-ci produise une action spécifique. Cependant, cette comparaison est imparfaite, car le contrôle du manipulateur ne porte pas directement sur l'action de sa victime (contrairement à l'ingénieur qui télécommande les gestes de sa machine), mais plutôt sur le type de structure motivationnelle qui assure l'accomplissement de l'action désirée par le manipulateur. En d'autres termes, le manipulateur ne constraint pas sa victime à agir, mais il la constraint à vouloir agir, ce qui donne lieu à des résultats similaires. Il existe, néanmoins, une différence centrale entre ces deux formes de contrainte. En effet, la victime manipulée croit être libre, contrairement au robot. Elle remplit même, dans ce cas précis, les conditions compatibilistes du libre arbitre (désir de second ordre et sensibilité aux raisons), mais demeure cependant manipulée. Grâce à ce type d'expériences de pensée, les incompatibilistes sont, dès lors, en possession de contre-exemples solides aux thèses de l'identification, dont le pouvoir intuitif repose principalement sur l'évidence selon laquelle la manipulation est incompatible avec la liberté :

Cas 1. Le professeur Plum, qui ressemble à un être humain ordinaire, a été créé par des neuroscientifiques qui le manipulent directement, grâce à l'utilisation d'une technologie de type radio. Supposons que ces neuroscientifiques agissent "localement" et dirigent les processus de raisonnement par lequel ses désirs sont provoqués et modifiés. Ils produisent chacun de ses états mentaux, seconde après seconde, en appuyant sur une série de boutons, de telle sorte que son processus de raisonnement soit rationnellement égoïste et qu'il finisse par tuer Mme White. Plum n'agit pas contrairement à son caractère puisqu'il est manipulé pour être rationnellement égoïste. Son désir effectif de premier ordre de tuer Mme White est conforme à ses désirs de second ordre. Par ailleurs, le processus de raisonnement de Plum illustre les différentes composantes de la sensibilité faible aux raisons, dans la mesure où s'il avait eu de bonnes raisons de faire autrement, il aurait fait autrement, dans certaines circonstances.

Intuition : Plum n'est pas responsable du meurtre de Mme White dans le Cas 1.

Cas 2. Plum est comme un être humain ordinaire, mais il a été créé par des neuroscientifiques de sorte qu'il soit rationnellement égoïste. Cette nature a été voulue par ses créateurs, qui, bien qu'ils ne le contrôlent pas directement, l'ont programmé pour qu'il décide d'être égoïste dans toutes les situations. Cette prédestination a pour résultat que, dans les circonstances dans lesquelles il se trouve maintenant, il est causalement déterminé, en vertu de son processus de sensibilité faible aux raisons et de la nature de ses désirs de premier et second ordre, à commettre le meurtre de Mme White.

Intuition : Si Plum n'est pas responsable du meurtre de Mme White dans le Cas 1, il n'est pas responsable dans le Cas 2 non plus.

Cas 3. *Plum est un être humain ordinaire, mais il a été formaté par un endoctrinement rigoureux voulu par sa famille et sa communauté religieuse, de sorte qu'il est souvent, mais pas exclusivement, rationnellement égoïste. Son formatage a eu lieu à un âge trop précoce pour qu'il ait désormais la capacité d'infléchir les pratiques qui ont déterminé son caractère permanent. Dans sa situation actuelle, Plum est faiblement sensible aux raisons et possède les désirs de premier et second ordre qui aboutissent au meurtre de Mme White.*

Intuition : Si Plum n'est pas responsable du meurtre de Mme White dans les Cas 1 et 2, il n'est pas responsable dans le Cas 3 non plus.

Dans cette suite, le transfert d'intuition repose en premier lieu sur le cas 1. Si Plum n'est pas responsable du meurtre de Mme White dans le premier cas, c'est parce qu'il est la marionnette des neuroscientifiques qui contrôlent ses motifs à chaque instant. Bien que Plum soit capable de s'identifier à ses actes de façon à satisfaire les conditions compatibilistes, l'identification est illusoire compte tenu de l'origine manipulée de ses états mentaux. Il paraît évident, donc, qu'une telle forme de manipulation n'est pas compatible avec la responsabilité morale.

Par suite, Plum n'est pas responsable du meurtre de Mme White dans le cas 2 non plus, car l'écart temporel qui le sépare de sa manipulation ne diminue pas l'effet de la manipulation elle-même. Plum est toujours la marionnette des neuroscientifiques qui l'ont programmé avant sa naissance, ce qui l'empêche d'agir librement.

Enfin, Plum n'est pas responsable du meurtre de Mme White dans le cas 3 non plus, car la seule différence entre le cas 2 et le cas 3 réside dans l'origine de la manipulation. Or, cette différence ne semble pas signifiante, en ce qu'une manipulation d'origine familiale n'est pas inférieure à celle des neuroscientifiques. Il paraît évident, en effet, que l'origine de la manipulation ne réduit pas l'effet de la manipulation elle-même. Dès lors, si un endoctrinement est à l'origine de la manipulation de Plum, il n'est pas responsable du meurtre de Mme White non plus.

Par ailleurs, Pereboom transfère cette intuition vers une situation de déterminisme ordinaire, dans le but de montrer que les cas de manipulation et de déterminisme partagent la même incompatibilité avec la responsabilité morale. Si les sujets manipulés ne sont pas libres, alors les sujets déterminés non plus :

Cas 4. *Le déterminisme est vrai et Plum est un être humain ordinaire, élevé dans des circonstances normales. Il est souvent, mais pas exclusivement, rationnellement égoïste. Le meurtre de Mme White par Plum résulte d'un processus de sensibilité faible aux raisons, et Plum possède des désirs de premier et de second ordre qui le motivent à agir de tel façon.*

Intuition : Si Plum n'est pas responsable du meurtre de Mme White dans les Cas 1,2 et 3, il n'est pas responsable dans le Cas 4 non plus.

Selon Pereboom, ce dernier mouvement est permis, dans la mesure où il n'existe pas de critère pertinent pour limiter l'intuition au cas 3 seulement. Pourquoi, en termes compatibilistes, Plum serait-il responsable du meurtre de Mme White dans le cas 4, alors qu'il ne l'est pas dans le cas 3 ?

En effet, il paraît plus prudent d'admettre que Plum n'est pas responsable du meurtre de Mme White dans le cas 4 non plus, car la détermination causale, comme dans les cas de manipulation, fait de ses actions le produit de facteurs qui échappent à son contrôle. Même si la détermination n'est plus dirigée vers une finalité prévue par un acteur extérieur, l'absence de cet acteur ne diminue pas, pour autant, la finalité de la détermination causale elle-même. Ainsi, la détermination externe est le facteur commun aux cas de manipulation et de déterminisme. Les quatre cas étudiés ont effectivement pour caractéristique commune qu'ils mettent en scène des situations où l'action est le résultat de facteurs qui échappent au contrôle de l'agent et c'est cette propriété qui invalide la responsabilité morale. Or, si cette propriété invalide la responsabilité morale dans les trois premiers cas, alors il doit également invalider la responsabilité morale dans le dernier cas. En conclusion, Plum est toujours déterminé à former les états mentaux qui aboutissent au choix de tuer Mme White et il n'est pas responsable de son acte :

"More generally, if an action results from a deterministic causal process that traces back to factors beyond the agent's control, then he is not morally responsible for it."⁴⁶

Pour plus de clarté, il est possible de formaliser l'idée de Pereboom en un argument de la manipulation simplifié :

1. Si un agent est manipulé, alors il n'est pas responsable de ses actes.
2. Si le déterminisme est vrai, alors il n'y a aucune différence pertinente entre un agent manipulé et un agent normal.
3. Donc, si le déterminisme est vrai, personne n'est moralement responsable.

La portée de cet argument est décisive, parce qu'elle contraint les compatibilistes à faire un choix dangereux pour leur position. En effet, la première prémissse est intouchable, dans la mesure où une théorie qui soutiendrait la responsabilité morale des victimes de manipulation tombe vraisemblablement dans l'absurde. Dès lors, les compatibilistes ont deux choix possibles : soit admettre la conclusion de l'argument (ce qui est inenvisageable pour un compatibiliste), soit proposer un critère discriminant qui explique en quoi les cas de manipulation sont pertinemment différents des cas déterministes, du point de vue de la responsabilité morale (il faut, dans ce cas, éviter une démarche *ad hoc*, peu convaincante, qui nuirait à la crédibilité théorique.)

⁴⁶ Pereboom, D. (2001). *Living Without Free Will*. Cambridge University Press.

Bien que je doute grandement des ressources compatibilistes en cette matière, il est peut-être nécessaire, pour le bien de la réflexion, de supposer qu'il soit possible de trouver un tel critère. Dans le meilleur des cas, ce critère devrait permettre d'identifier une propriété présente dans la manipulation, mais faisant défaut au déterminisme. Ainsi, le compatibiliste serait en mesure de proposer que cette propriété seulement invalide la responsabilité morale.

À ce stade, je crois qu'il est dans l'intérêt des incompatibilistes de mettre fin à ces espoirs, en développant un argument de la manipulation qui donne, si possible, encore plus de crédibilité à la prémissse 2. Ici, le but des incompatibilistes est de mettre au point un nouvel argument de la manipulation qui donne lieu à des intuitions si flagrantes que la recherche de ce critère semblera directement compromise. Il existe un argument que l'on doit à Alfred Mele⁴⁷ et qui atteint, je crois, cet objectif. L'argument du Zygote consiste en deux étapes et a pour avantage de soutenir la prémissse 2 avec plus de clarté et de simplicité que ne le permettaient les cas de Pereboom, tout en conservant, cependant, la même structure essentielle. Dans cette manipulation, ce ne sont pas des neuroscientifiques qui contrôlent une victime, mais une déesse omnipotente évoluant dans un monde déterministe, Diana, qui crée un zygote, dans le but qu'Ernie, soit la personne issue du zygote, commette un meurtre trente ans plus tard :

Diana crée un zygote Z dans Mary. Elle combine les atomes de Z, parce qu'elle veut qu'un certain événement E se produise trente ans plus tard (p.ex. un meurtre). De sa connaissance de l'état de l'univers juste avant qu'elle ne crée Z et des lois de la nature de son univers déterministe, elle déduit qu'un zygote avec précisément la constitution de Z, situé en Marie, se développera en un agent autonome qui, dans trente ans, jugera, sur la base d'une délibération rationnelle, qu'il est préférable de faire A et fera A, sur la base de ce jugement, en entraînant ainsi l'évènement E. [...] Trente ans plus tard, Ernie est une personne en bonne santé mentale, autonome, qui exerce régulièrement ses pouvoirs de maîtrise de soi et n'a aucune attitude pertinente, forcée ou produite de manière coercitive. De plus, ses convictions sont propices à une délibération éclairée sur toutes les questions qui le concernent, et il est un délibérateur fiable. Il satisfait donc les conditions compatibilistes pour avoir librement fait A.⁴⁸

Dans ce monde relativement similaire au nôtre, bien qu'Ernie satisfasse les conditions compatibilistes, il semble intuitivement clair qu'il n'agit pas librement et qu'il n'est pas responsable de l'évènement, car Ernie a précisément été créé pour qu'il réalise l'évènement de façon certaine. Dès lors, sa capacité à s'identifier à l'action ne semble pas, dans un tel cas, suffisante pour la responsabilité morale. En effet, serait-il juste de le punir alors qu'il manquait évidemment de contrôle sur un meurtre inévitable dès l'origine ? La réponse semble très intuitivement pencher pour la négative.

À présent, considérons Bert, le double contrefactuel d'Ernie, atome pour atome. Bert est un être humain ordinaire en tous points de vue, partageant exactement les mêmes caractéristiques et la

⁴⁷ Mele, A. R. *Free Will and Luck*, New York: Oxford University Press. doi:10.1093/0195305043.001.0001

⁴⁸ Traduction adaptée de Mele, A.R. *Free Will and Luck*, New York: Oxford University Press. doi:10.1093/0195305043.001.0001

même histoire qu'Ernie (évoluant dans un monde déterministe, fils de Mary, rationnel, autonome, commettant le même meurtre à 30 ans), à la seule différence que Bert n'a pas été créé par Diana. La réalité de Bert est donc identique, point par point, à celle qui est décrite dans l'expérience de pensée d'Ernie, si ce n'est que son zygote a été produit naturellement.

Selon Mele, dans la mesure où Ernie et Bert sont la même personne en tout point, il n'existe pas de raison pertinente pour que Bert soit responsable du meurtre, alors qu'Ernie ne l'est pas. Dès lors, s'il n'y a pas de différence moralement pertinente entre la situation de manipulation et celle de déterminisme, alors il faut en conclure que Bert, non plus, n'est pas moralement responsable. Ici, cette conclusion est sans doute facilitée par le caractère très négligeable de l'unique différence qu'il existe entre les deux contrefactuels. En effet, dans le monde de Bert, la détermination divine est simplement substituée par une détermination d'ordre naturel ou biologique. Mais cette substitution modifie-t-elle cependant nos intuitions à propos du caractère déterminé de son action ?

Si la réponse est non, comme je le crois, la ressemblance entre les deux mondes semble telle qu'il est, d'emblée, impossible de chercher un critère de différentiation qui sauverait les compatibilistes. Dès lors, le succès de l'argument est large : il faut en conclure finalement que personne n'est véritablement responsable dans un contexte déterministe.

Face à cet argument de la manipulation, il subsiste, je crois, une réponse compatibiliste à considérer. En effet, un compatibiliste pourrait, ici, être tenté de proposer de reprendre l'argument incompatibiliste, mot à mot, afin de le retourner contre lui-même (voir un bon exemple de cette stratégie chez Mckenna⁴⁹). En effet, s'il n'y a aucune différence moralement pertinente entre Ernie et Bert, pourquoi devrions-nous partir de la non-liberté d'Ernie pour prouver celle de Bert, plutôt que l'inverse, c'est-à-dire de la liberté de Bert pour prouver celle d'Ernie. Dans la mesure où la non-liberté d'Ernie repose essentiellement sur notre soutien intuitif, les compatibilistes peuvent encore soutenir que la liberté de Bert repose, elle aussi, sur notre soutien intuitif. Ainsi, s'il n'y a aucune différence moralement pertinente entre Ernie et Bert, il est possible, ensuite, de transférer cette intuition vers le cas de manipulation pour prouver l'inverse des incompatibilistes, à savoir que Bert et Ernie agissent librement et sont moralement responsables.

Toutefois, peut-on réellement soutenir que ces deux soutiens intuitifs se valent ? Peut-on véritablement soutenir que l'intuition de la non-liberté d'un agent manipulé vaille celle de la liberté d'un agent déterminé ? Il me semble très clairement que non. En fin de compte, je crois que nos réponses doivent reposer sur une honnêteté philosophique, sans laquelle la réflexion intuitive devient, elle-même, impossible. En effet, en défendant que l'intuition de la liberté de Bert vaut celle

⁴⁹ McKenna, M. *A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument*. *Philosophy and Phenomenological Research*, vol. 77, no. 1, 2008, pp. 142–59. JSTOR, <http://www.jstor.org/stable/40041222>.

de la non-liberté d’Ernie, les compatibilistes n’offrent aucun argument supplémentaire à leur position, mais jettent plutôt un discrédit fondamental sur le pouvoir de l’intuition elle-même, ce qui empêche le débat. Or, il est à noter que ce discrédit ne s’attaque pas uniquement aux incompatibilistes, mais rejaillit aussi sur le compatibilisme, ainsi qu’une bonne partie de la pratique philosophique elle-même. Dès lors, je crois qu’il est nécessaire de veiller à éviter un déclin dialectique qui n’est bénéfique nulle part. En effet, les intuitions sont nécessaires au débat et doivent être critiquées lorsqu’elles sont incorrectes et acceptées lorsqu’elles sont correctes. En ce qui concerne les arguments de la manipulation, ce principe s’applique également. Aussi, je ne suis pas opposé à une potentielle réponse compatibiliste qui viserait à expliquer en quoi les intuitions incompatibilistes sont infondées. Je m’oppose toutefois à l’idée que l’intuition est toujours douteuse. Dès lors, je considère l’apport intuitif des arguments de la manipulation comme un avantage théorique, plutôt qu’une faiblesse. De ce point de vue, il me semble que l’argument est véritablement décisif. En effet, les cas que nous avons étudiés constituent une très sérieuse objection au compatibilisme, en mettant en cause les fondements de la théorie. Grâce aux arguments de la manipulation, il nous apparaît clairement que le libre arbitre est un concept métaphysique plutôt que psychologique. Grâce aux contre-exemples, nous comprenons que l’identification n’est pas suffisante pour la responsabilité morale. Être libre se conçoit, donc, en termes de possibilités, plutôt qu’en termes d’identification. Or, si le déterminisme est vrai, la seule possibilité est actuelle et nous devons admettre que cette thèse est incompatible avec le libre arbitre.

2.5 Libertarianisme

Je viens de proposer que le compatibilisme se heurte à des arguments décisifs. Si ma ligne a été assez convaincante, il faut en conclure que le déterminisme et le libre arbitre sont incompatibles. Cependant, cette première réponse n'est pas suffisante pour poser définitivement l'impossibilité du libre arbitre. Comme nous l'avons vu, il existe encore deux versions incompatibilistes qui défendent l'existence du libre arbitre, en niant celle du déterminisme. Ces thèses garantissent l'existence du libre arbitre, en employant des moyens différents du compatibilisme. Il me faut alors les considérer également avec attention. Plus précisément, le libertarianisme ne nie pas l'existence de mécaniques déterministes dans la nature, mais réfute le déterminisme à l'échelle de la délibération humaine. Ainsi, lors de la délibération, la conjonction des lois naturelles et du passé n'implique pas un seul et unique futur possible. Les libertariens refusent donc le déterminisme, *par définition*, au profit du libre arbitre et de la responsabilité morale. Puisque le libre arbitre est incompatible avec le déterminisme, il faut donc également postuler la fausseté du déterminisme. Dans ces conditions, les libertariens prétendent garantir la responsabilité morale.

En adoptant cette perspective, toute la difficulté consistera, cependant, dans l'explicitation des raisons qui rendent le déterminisme inefficace à échelle humaine, alors qu'il est probablement vrai à l'échelle des choses qui nous entourent. Ces raisons doivent, par ailleurs, être assez pertinentes pour justifier l'abandon du principe d'économie selon lequel une hypothèse n'a pas de valeur, si elle n'a pas de pouvoir explicatif supérieur à la théorie dominante. Or, puisque le déterminisme décrit relativement bien la causalité naturelle, il faut trouver une raison significative pour motiver l'opposition entre la causalité déterministe régissant le monde d'un côté, et la causalité humaine de l'autre. Il n'est cependant pas clair que les grands intérêts moraux que nous voulons défendre remplissent une fonction rationnelle pour autant. Dans les développements qui suivent, je présente les deux formes de libertarianisme en essayant de mettre l'accent sur les conditions du libre arbitre qu'elles prétendent satisfaire. J'aborderai la causalité agentive, dans un premier temps, et la causalité événementielle, ensuite.

2.5.1 Libertarianisme de la Causalité Agentive

Selon le libertarianisme de la causalité agentive, les actions humaines ne sont pas le résultat d'évènements antérieurs, déterminés par des facteurs externes, mais sont directement produites par les agents eux-mêmes. Lorsque nous disons, par exemple, « Léa a délibérément assassiné son voisin. », nous ne presupposons pas que les différents évènements psychologiques liés à sa personne (dispositions, croyances, désirs, etc...) ont causé l'assassinat, mais plutôt que Léa, elle-même, en tant que substance, a directement causé la mort de son voisin. Ainsi, le libertarianisme de la causalité agentive correspond étroitement à la formulation littérale de nos énoncés quotidiens. En effet, dans nos descriptions ordinaires de l'agentivité, nous lions causablement l'action à l'agent de façon directe. Par ailleurs, les libertariens conçoivent les agents comme des causes libres et non-déterminées des évènements qu'ils produisent. Bien que tout évènement soit toujours déterminé par des facteurs antérieurs, les agents, en tant que substances, sont, quant à eux, conçus comme fondamentalement libres.

Il est difficile de définir avec plus de précision la forme d'un modèle standard de causalité agentive, dans la mesure où ses principaux représentants ont souvent postulé l'existence du libre arbitre, sans clarifier d'abord la structure métaphysique qu'il doit satisfaire. En effet, la position semble reposer, avant tout, sur un sens commun considérable, plutôt qu'un modèle théorique. À ce titre, il est probable que le crédit accordé à la causalité agentive tienne en grande partie au grand soutien intuitif que représente notre expérience de l'action, à la première personne. En effet, du point de vue de l'expérience phénoménale, il semble que nos actions sont produites librement, car elles nous apparaissent comme s'il était possible d'agir autrement et comme si nous en étions la source ultime.⁵⁰ Ainsi, la causalité agentive satisfait entièrement les conditions du libre arbitre que nous avions identifiées dans la conception traditionnelle. Puisque les agents bénéficient d'un contrôle radical sur leurs actions, ils en constituent la source d'une telle manière qu'ils auraient pu réellement faire autrement. Ils sont capables d'initier causablement une action dont la réalisation dépend uniquement d'eux-mêmes et créent, par leur volonté libre, des chaînes causales alternatives. Dès lors, si toute action est fondamentalement initiée indépendamment des évènements antérieurs, les agents possèdent une capacité robuste d'agir autrement et sont la source *ultime* de leurs actions.

⁵⁰ Il n'est cependant pas clair que nos intuitions soient si tranchées en cette matière. (voir Nahmias, E. Morris, S., Nadelhoffer, T. et Turner, J. (2004). *The Phenomenology of Free Will*. Journal of Consciousness Studies, 11(7-8), 162–179.)

2.5.2 Libertarianisme de la Causalité Evènementielle

Selon le libertarianisme de la causalité évènementielle, ce sont les événements mentaux associés aux agents, par exemple leurs émotions, leurs désirs ou leurs croyances, qui causent les actions. La causalité évènementielle est donc similaire à un modèle déterministe standard, à la seule différence qu'elle permet la possibilité qu'un évènement en cause un autre, de façon indéterministe (ce qui ne signifie pas que tous les événements sont toujours causés de manière indéterministe, mais seulement que certains le sont). Il est à noter, tout d'abord, que l'idée même d'une causalité indéterministe n'est pas absurde, compte tenu de certaines interprétations de la mécanique quantique, promues en physique contemporaine. S'il existe effectivement un indéterminisme fondamental dans la réalité subatomique des choses, alors il n'est pas interdit de concevoir l'hypothèse qu'un indéterminisme infra-neuronale, aux effets macroscopiques, soit également possible, lorsque nous délibérons. Sur cette base, il est ensuite possible de concevoir nos choix comme étant le résultat d'une causalité évènementielle indéterministe : nos raisons causeraient de façon indéterministe notre prise de décision, qui causerait ensuite l'action de façon déterministe. Par ailleurs, l'introduction de l'indéterminisme au sein de la causalité suffit pour contredire la thèse du déterminisme universel et permet d'entrevoir, sur cette base, la possibilité d'un libre arbitre. En effet, si nos prises de décision ne sont pas déterminées par les événements antérieurs et qu'il est, *a priori*, impossible de prédire l'action qui sera actualisée, il peut sembler désormais que nous ayons davantage de contrôle agentif. Selon les libertariens, cette modification centrale a précisément pour but de permettre de sauver le libre arbitre et la responsabilité morale. En effet, s'il existe un indéterminisme au niveau de nos prises de décision, alors il est métaphysiquement possible que l'évènement actuellement produit n'ait jamais eu lieu et qu'un autre évènement fût produit à la place. La première condition du libre arbitre est alors remplie. Néanmoins, concernant la seconde condition, il n'est pas clair que le fait de postuler un indéterminisme entre les événements permette aux agents d'être la source ultime de leurs actions. En effet, si nos décisions sont causalement indéterminées, il apparaît alors que ce ne sont pas les agents, mais le hasard qui détermine l'action en dernier lieu. Cette introduction de l'indéterminisme est particulièrement problématique, en ce qu'elle ne semble pas augmenter notre contrôle sur les actions. Ainsi, si les décisions que nous prenons sont finalement réduites à des occurrences purement hasardeuses, les libertariens auront sauvé le principe des possibilités alternatives à un prix qui n'est pas envisageable, car le principe à lui seul n'est pas suffisant pour le libre arbitre.

Pour cette raison, il semble encore nécessaire aux libertariens d'expliquer comment l'agent occupe une place déterminante dans la réalisation de ces différentes possibilités, sans quoi le contrôle est

perdu. Dans l'objectif de combler cette lacune, Robert Kane⁵¹ donne une réponse similaire à celle des compatibilistes et propose de lier nos raisons et nos actions par une forme d'identification. Selon sa théorie, nos raisons ne causent pas systématiquement nos choix sur un mode indéterministe, mais seulement lorsque nous nous trouvons dans une situation où nous avons autant de raisons nous poussant à agir dans deux directions contradictoires (ou plus). Dans un tel dilemme, nous avons toujours de bonnes raisons d'agir pour les deux possibilités d'actions soumises à l'indétermination causale. Ces raisons permettent ensuite l'identification, parce qu'elles se rattachent directement aux différents états mentaux qui constituent notre nature. Nous nous identifierons ainsi à l'action actuelle, car le lien entre les raisons et l'action est sauvegardé dans les deux cas.

Par conséquent, l'indéterminisme n'est pas à concevoir, selon Kane, tel un hasard vicieux, mais plutôt comme une incertitude fondamentale quant à l'actualisation de différentes options fidèles au caractère d'un agent. En somme, son modèle a pour avantage, non seulement de satisfaire la condition de source agentive, mais encore de garantir simultanément le principe des possibilités alternatives, auquel les compatibilistes étaient contraints de renoncer.

⁵¹ Kane, R. (1996), *The Significance of Free Will*. New York: Oxford University Press.

2.6 Arguments contre le Libertarianisme

Dans cette section, je propose des objections d'ordre conceptuel aux deux positions que je viens de présenter. Je ne formulerais pas d'objections empiriques⁵², car il me semble que la première catégorie de critiques remplit efficacement sa fonction philosophique et que la deuxième n'est utile qu'en second recours. Par ailleurs, en attaquant la cohérence même des théories, les objections conceptuelles bénéficient certainement d'un pouvoir de conviction plus radical.

Tout d'abord, j'analyserai l'incohérence de la notion de *causa sui*, que le libertarianisme de la causalité agentive abrite en lui-même. Je montrerai alors que ce concept mène à une régression vicieuse. Dans un second temps, je soutiendrai que le modèle de la causalité évènementielle, même dans sa forme la plus épurée, est incapable de se défaire de la présence d'un hasard incompatible avec le libre arbitre et la responsabilité morale.

2.6.1 *Causa Sui*

Pour mettre en lumière l'incohérence de la causalité agentive, reprenons d'abord l'exemple de Léa qui assassine délibérément son voisin. D'après les libertariens, Léa commet cet assassinat parce qu'elle a décidé librement de le faire, à la suite d'un processus indéterministe. Pour autant, il ne faut pas comprendre l'indétermination en question selon l'idée que Léa aurait agi de façon hasardeuse et sans raison, parce qu'elle ne serait alors ni libre, ni responsable de son acte. En effet, selon les libertariens de la causalité agentive, lorsque nous disons que Léa n'était pas déterminée à assassiner son voisin, il faut plutôt comprendre qu'elle n'était pas déterminée à le faire par des facteurs extérieurs à elle-même. À ce titre, les actions de Léa ne sont pas indéterminées, mais *autodéterminées*, ce qui signifie qu'elle agit librement, en considérant les différents motifs qui lui sont propres. Avant le meurtre, on peut imaginer, par exemple, que Léa jugeait l'attitude suspecte de son voisin comme une menace ou alors qu'elle nourrissait, depuis toujours, des désirs inavouables de violence qui l'ont poussée à agir.

Cependant, les libertariens ne peuvent pas admettre non plus, que les raisons qui ont motivé l'assassinat reposaient sur des facteurs qui échappaient au contrôle de Léa (sa nature paranoïaque ou son sadisme, par exemple). Si tel était le cas, il faudrait en conclure également qu'elle manquait de contrôle sur son acte, comme nous l'avons montré à l'occasion des objections au compatibilisme.

⁵² Il est en effet possible de formuler de nombreuses objections au libertarianisme sur la base de son invraisemblance physique.

Les libertariens doivent donc défendre, par ailleurs, que Léa disposait d'un contrôle sur ses motivations, qu'elle les a choisies et qu'elles n'ont pas été simplement subies. À ce stade, il paraît clair, d'ores et déjà, que la causalité agentive se situe dans un intervalle délicat. En effet, si les raisons de Léa ne déterminent pas assez ses actions, alors elles sont hasardeuses. Mais si les raisons de Léa déterminent trop ses actions, alors elles sont inéluctables. Pour sortir de cette tension, les libertariens de la causalité agentive soutiennent que les agents libres possèdent également la capacité de contrôler leurs motifs et les raisons qui les poussent à agir. Léa n'est pas seulement la cause libre de ses actions, mais également la cause libre d'elle-même (*causa sui*), dans la mesure où elle choisit les traits et caractéristiques qui la constituent sur le plan psychologique : Léa a assassiné son voisin librement, parce qu'elle a choisi d'être qui elle est. Si elle avait décidé d'être autrement, c'est-à-dire d'avoir d'autres croyances ou désirs, alors elle n'aurait pas commis ce meurtre.

Toutefois, ce point pose un problème. En effet, la *causa sui* semble contenir une incohérence très problématique, mise en lumière par Galen Strawson.⁵³ Pour que Léa soit responsable de son assassinat, nous savons désormais qu'il faut également qu'elle soit responsable du caractère malveillant à partir duquel s'est fondée l'action. Mais pour être responsable de son caractère malveillant, il faut l'avoir choisi. Or, il faut postuler que Léa préexistait à ce caractère, puisqu'il ne peut pas exister de choix sans une agentivité qui l'opère et c'est sur la base de cette nature fondamentale que Léa a choisi son caractère. Mais à nouveau, cette nature préexistante doit avoir des raisons pour que Léa choisisse son caractère librement, sinon c'est le hasard qui en est à l'origine. Ainsi, pour être vraiment libre, il ne suffit pas que Léa décide de son acte et de son caractère, il faut également qu'elle décide de la nature antérieure à partir de laquelle elle a décidé de son caractère et ainsi de suite jusqu'à l'infini. Selon Galen Strawson, les implications de la causalité agentive démontrent finalement qu'elle est impossible, en raison de la réduction générée par le concept même. En effet, la responsabilité morale implique une autodétermination quasi-divine qui ne peut raisonnablement être attendue d'un être humain :

« If we are responsible (...) then we have a prerogative which some would attribute only to God: each of us, when we act, is a prime mover un-moved⁵⁴

Pour cette raison, le libertarianisme de la causalité agentive apparaît comme une position caduque, tant que les libertariens restent incapables de résoudre le problème conceptuel. Cela étant dit, on

⁵³ Strawson, G. (1994). *The Impossibility of Moral Responsibility*. Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 75(1/2), 5–24. <http://www.jstor.org/stable/4320507>.

⁵⁴ Chisholm, R. (1964). *Human Freedom and the Self*. In Robert Kane (ed.), *Free Will*. Blackwell.

conçoit mal comment la réduction à l'infini pourrait être évitée à partir du moment où la responsabilité morale semble requérir un libre arbitre au sens fort.

2.6.2 Hasard

Le principal problème conceptuel de la causalité évènementielle concerne le hasard, comme je l'ai brièvement mentionné lors de la présentation. En effet, la substitution du déterminisme par l'indéterminisme ne semble pas être une solution viable à première vue, puisque le hasard semble également incompatible avec le libre arbitre. Robert Kane, à qui nous devons certainement la théorie libertarienne la plus aboutie, prétend néanmoins qu'on peut se prémunir des méfaits du hasard en ajustant le modèle de la causalité évènementielle. Selon sa thèse, le libre arbitre existe, sitôt que nos actions sont causées par un processus indéterministe intervenant dans notre structure cérébrale. En effet, quand un agent est dans un tel dilemme qu'il est incapable de trancher entre des raisons équivalentes, l'action produite de façon indéterministe falsifie l'inéluctabilité du futur, sans être hasardeuse pour autant, car elle reste fidèle à ses raisons. On peut donc concevoir la position de Kane comme un intermédiaire fluide, tirant avantage du compatibilisme pour garantir l'identification et de l'indéterminisme pour sauver la possibilité alternative. Je pense néanmoins que cette théorie hérite également des inconvénients inhérents à ses avantages. Ainsi, il est possible de critiquer la théorie de deux façons. Tout d'abord, si le processus qui transforme les raisons en options est déterministe, comme le prétend Kane (dans l'espoir de sauver l'identification et donc la source agentive), la position est susceptible de devoir faire face aux mêmes objections que celles que nous avions formulées à l'encontre du compatibilisme. En effet, puisque les options disponibles sont déterminées par des facteurs qui échappent à notre contrôle, l'action le sera également, même si elle est indéterminée, dans un second temps. Imaginons qu'un agent hésite entre deux actions immorales (tricher à un examen en utilisant son portable et tricher en utilisant des fiches de papier), de telle sorte que le choix final soit indéterminé. Cet agent bénéficie de la possibilité robuste d'agir autrement car les deux actions sont indéterminées, mais il n'a cependant pas la possibilité pertinente de ne pas tricher, en raison de la nature même de ses raisons qui sont hors de contrôle. Il serait étrange de défendre que l'indéterminisme permette la responsabilité morale dans cette situation, puisque l'agent reste déterminé à agir immoralement dans les deux cas.

La seconde critique porte sur le hasard. Bien que le lien entre nos raisons et les différentes options que nous envisageons ne soient effectivement pas hasardeux, cela signifie-t-il pour autant que le choix parmi les deux options ne l'est pas non plus ? Il y a ici une confusion à éviter. Si les deux options possibles correspondent bel et bien à la nature de l'agent et que celui-ci s'identifie dans tous les cas,

il semble néanmoins que l'agent n'opère pas un rôle suffisant dans la prise de décision finale, ce qui cède toujours une place au hasard. En d'autres termes, dans le modèle de Kane, l'agent bénéficie effectivement d'un certain contrôle sur ses options, sans avoir, pour autant, de contrôle au niveau du processus indéterministe qui en actualisera l'une des deux. La place occupée par le hasard, dans cette conception du libre arbitre, pourrait ensuite entraîner des conséquences inacceptables au niveau de nos pratiques morales.

Alfred Mele⁵⁵ propose une expérience de pensée qui met en évidence ce problème. Imaginons que Mike soit tiraillé entre deux actions possibles, l'une étant immorale et l'autre morale (rendre ou non un portefeuille trouvé). Puisqu'il est incapable de choisir entre les deux options qui sont compatibles avec l'ensemble de ses raisons morales et égoïstes, un processus indéterministe s'initie dans son cerveau et Mike choisit finalement de rendre le portefeuille. Imaginons maintenant une situation contrefactuelle identique où l'indéterminisme aboutit au choix contraire, c'est-à-dire au fait de garder le portefeuille. Selon Kane, Mike a agi librement dans les deux cas, parce qu'il avait la possibilité robuste d'agir autrement et parce que les deux actions sont cohérentes avec la personne qu'il est. En conséquence, il faut le blâmer dans le cas où il garde le portefeuille, et le louer dans le cas où il le rend. Pourtant, il semble, en un sens moral, que la prise de décision finale échappe totalement au contrôle de Mike, dans les deux cas. L'action demeure hasardeuse dans la mesure où ses raisons, bien qu'elles expliquent les options disponibles, n'expliquent pas, cependant, pourquoi il rend le portefeuille dans le premier cas, et le garde dans le second. Ainsi, l'action libre est toujours hasardeuse en dernière instance. Les deux issues métaphysiquement possibles reposant autant sur le hasard, il semble étrange qu'elles justifient ensuite des évaluations morales contraire. Lorsque nous blâmons une mauvaise action, nous voulons que l'évaluation puisse rejoindre sur les motifs ou le caractère de l'agent. Or, dans le modèle de Kane, un même caractère et de mêmes motifs sont susceptibles d'être hasardeusement bons ou mauvais, l'agent étant comme spectateur de ses actions en dernière instance. Par conséquent, la causalité évènementielle réduit les bonnes actions à des actions chanceuses et nous ne sommes pas près de nous contenter de cette conclusion.⁵⁶ Dès lors, le type de contrôle permis par cette théorie est encore insuffisant pour la responsabilité morale. Plus encore, le problème du hasard est probablement irréductible, dès lors que l'on postule une causalité évènementielle indéterministe.

⁵⁵ Mele, A. R. (2006). *Free Will and Luck*. New York, US: Oxford University Press.

⁵⁶ Levy, N. (2011). *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford, GB: Oxford University Press UK.

2.7 Du Scepticisme au Déterminisme Dur

Depuis le début de ce chapitre, je n'ai pas encore pris parti pour une position propre et me suis contenté d'une évaluation objective des théories du libre arbitre. Pour parachever cette réflexion, il me faut désormais proposer une interprétation générale des conclusions que j'ai tirées. Tout d'abord, je crois avoir montré de façon satisfaisante que la position compatibiliste n'était pas raisonnablement tenable, du moins dans les formes ambitieuses que j'ai discutées. Je pense, en réalité, que la grande popularité dont elle jouit dans la communauté philosophique tient d'avantage aux intérêts qu'elle permet de défendre, qu'à sa rationalité propre. En effet, parvenir à la conclusion selon laquelle la responsabilité morale existe, qu'importe la vérité du déterminisme, représente des avantages pratiques qui ne sont pas remis en question. Cependant, le maintien de ces intérêts se fait, à mon sens, au prix d'un échec théorique trop manifeste. Après analyse, les thèses de l'identification aboutissent, en effet, à une conception affaiblie du libre arbitre qui s'avère parfois en telle contradiction avec nos intuitions qu'elle n'en garde que le nom. Je tiens à préciser, néanmoins, que je ne suis pas foncièrement opposé à l'usage d'une version faible du libre arbitre. Au contraire, il me semble que celle-ci pourrait occuper une fonction pragmatique qui nous permettrait, en pratique, de faire la différence entre une action qui suit une longue délibération et une action produite sous la menace d'une arme, par exemple. En revanche, je suis très opposé à l'idée centrale selon laquelle ce libre arbitre affaibli serait suffisant pour la responsabilité morale. En effet, il est possible qu'un certain degré de contrôle soit compatible avec le déterminisme, sans être nécessairement suffisant pour la responsabilité morale. Or, étant donné que nous avons expressément défini le libre arbitre en ces termes, le contrôle requis pour la responsabilité morale reste incompatible avec le déterminisme. Puisque nous avons un attachement primitif au libre arbitre et qu'il est incompatible avec le déterminisme, nous pouvons encore se demander si le déterminisme est faux. J'ai, dans un deuxième temps, appréhendé deux manières de rendre compte de la fausseté du déterminisme pour affirmer le libre arbitre. En ce qui me concerne, j'ai une sympathie pour la thèse libertarienne de la causalité agentive, principalement pour deux raisons. La première concerne le soutien phénoménologique dont elle semble bénéficier grandement et que je n'ai fait qu'esquisser dans mes réflexions préalables. En effet, lorsque nous expérimentons la délibération à la première personne, je crois que nous nous représentons le monde exactement tel que les libertariens décrivent l'agentivité humaine. La deuxième raison est que, de toutes les théories du libre arbitre, l'analyse ambitieuse promue par la causalité agentive est la seule qui soit vraiment suffisante pour la responsabilité morale à mon sens.

Néanmoins, à contre cœur, je crois qu'il est nécessaire d'abandonner cette position pour la seule raison qu'elle repose, comme je l'ai montré, sur une incohérence conceptuelle très sérieuse. Je considère, donc, le libre arbitre comme impossible, à moins que l'on réussisse à en fournir une analyse conceptuelle convaincante qui résisterait à la régression à l'infini identifiée. Bien que l'argument de Galen Strawson me semble particulièrement efficace, je ne suis pas certain de l'impossibilité *a priori* d'une telle entreprise. J'adopterais volontiers le libertarianisme dans ces conditions, qui ne sont malheureusement pas réunies à présent.

À ce stade, j'ai déjà refusé deux positions sur les quatre existantes (compatibilisme, causalité agentive, causalité évènementielle et déterminisme dur). Il me reste donc à trancher entre le libertarianisme de la causalité évènementielle et le déterminisme dur. Dans la mesure où l'indéterminisme postulé par les libertariens de la causalité évènementielle entraîne un problème de hasard incompatible avec la responsabilité morale, j'estime que la position échoue à expliquer convenablement en quoi l'introduction de l'indéterminisme causal est salutaire.

Par conséquent, il me reste à choisir entre, d'une part un modèle indéterministe incompatible avec le libre arbitre, et d'autre part un modèle déterministe également incompatible avec le libre arbitre. À ce titre, étant donné que le libertarianisme de la causalité évènementielle n'atteint pas l'objectif qu'il s'était fixé, il me semble que l'opposition entre la causalité déterministe universelle d'un côté et la causalité indéterministe des êtres humains de l'autre, devient superflue, en raison du principe d'économie que j'ai mentionné. Par un raisonnement abductif qui fait du déterminisme dur la meilleure explication (jusqu'à preuve du contraire), je rejoins finalement cette thèse. En résumé, ma position est essentiellement le résultat d'un raisonnement négatif qui tient compte des échecs des différentes théories. Je pense donc qu'il est possible d'être déterministe, sans nécessairement chercher à établir de preuves positives de l'universalité du déterminisme causal (ce qui est d'ailleurs difficilement envisageable). En ce sens, je ne partage pas, non plus, l'approche des déterministes classiques qui défendent, à l'image de Spinoza⁵⁷, une vision du monde positivement déterministe. Au contraire, le déterminisme auquel je souscris est avant tout le constat d'un échec, celui de la causalité agentive en premier lieu, et celui des autres théories dans un second lieu.

⁵⁷ Spinoza, B. *Éthique*. Édition établie, présentée et annotée par Charles Ramond. Collection Folio Essais, Gallimard, 2010.

Chapitre III : Le Problème des Attitudes Réactives

Dans ce dernier chapitre, je m'intéresse à un problème pratique qui contrarie le scepticisme à propos de la responsabilité morale. Dans le Chapitre I, j'ai défini les attitudes réactives comme un ensemble d'émotions qui impliquent une attribution de responsabilité morale. Or, si la responsabilité morale est impossible, pour les différentes raisons que j'ai données, il faut alors en conclure que nos attitudes réactives sont toujours inadéquates (*unfitting*) et doivent finalement être abolies ou révisées. Cependant, même les déterministes les plus convaincus doivent admettre que cette conclusion les met dans une position délicate, propre au scepticisme. Comment expliquer, en effet, la difficulté des sceptiques à appliquer leurs croyances ?

Il existe ici une étrange incohérence entre les résultats théoriques auxquels nous parvenons et l'ensemble des attitudes et comportements que nous adoptons pratiquement. Stephen Hawking raconte, dans « Une brève histoire du temps », une anecdote qui souligne avec ironie le problème dont je parle. Le récit est celui d'un physicien, déterministe convaincu, pour qui un seul futur est possible, compte tenu des lois de la nature. Le physicien argumente avec son ami, alors qu'ils se promènent dans la rue. Lorsque les deux hommes arrivent au bord d'une route, le scientifique s'interrompt pour regarder à droite, puis à gauche, avant de traverser, ce qui a pour effet de faire sourire son ami. L'ami sourit parce que le comportement pratique du physicien laisse entendre que différents futurs sont possibles, alors qu'il développait à l'instant un discours inverse.

Dans ce chapitre, je chercherai à montrer qu'il existe au sujet des attitudes réactives un problème similaire, qui tient à l'incapacité de se défaire d'une pratique que l'on croit infondée, mais vers laquelle nous sommes naturellement portés.

Dans un premier temps, j'identifierai les problèmes du déterminisme en prenant soin d'évacuer d'abord les faux-problèmes (3.1.), à savoir ceux qui n'ont pas de réel fondement. Parmi d'autres vrais problèmes (3.1.1 et 3.1.2), je défendrai que le problème le plus sérieux est celui du traitement des attitudes réactives (3.1.3). Ensuite, j'essayerai de proposer et discuter les solutions possibles au problème identifié. Je commencerai par introduire les thèses défendues par Strawson dans *Freedom and Resentment* (3.2.). Je distinguerai alors deux thèses fortes : la thèse naturaliste (3.2.1 et 3.2.2) et la thèse normative (3.2.3) et évaluerai dans quelle mesure elles résolvent le problème (3.2.4).

À la suite de cela, je formulerais des objections aux deux thèses strawsonniennes (3.3 et 3.4).

Ces objections me mèneront, enfin, à la défense d'une autre thèse, l'illusionnisme, qui est capable de résoudre le problème (3.5).

3.1 Identification des Problèmes du Déterminisme Dur

Pour parvenir au problème des attitudes réactives, ma méthode procèdera, en premier lieu, par l'élimination des critiques infondées que l'on peut formuler à l'encontre du déterminisme *prima facie*. Je commence en effet par écarter les problèmes que le déterminisme n'a pas à résoudre, pour parvenir dans un second temps, à ceux qui me paraissent les plus sérieux.

Tout d'abord, il existe une confusion entre fatalisme et déterminisme, à l'origine d'une partie de ce que j'appellerai les « faux-problèmes » du déterminisme dur. En effet, la thèse selon laquelle un seul futur est possible, *peu importe les choix que nous faisons*, diffère de la thèse selon laquelle un seul futur est possible, *en fonction des choix que nous faisons*. On saisit mieux cette différence à l'aune d'un exemple littéraire, car le fatalisme a souvent été mis en scène explicitement dans les tragédies grecques. Je présenterai d'abord un exemple de fatalisme, en le comparant ensuite au déterminisme. Considérons Œdipe, fils du roi de Thèbes, voué à un destin tragique avant sa naissance : une prophétie annonce que l'enfant tuera son père et épousera sa mère. Malgré les efforts de ses parents pour éviter la prophétie en l'abandonnant dans les bois, Œdipe est adopté et, bien des années plus tard, ignorant tout de son origine, finira par tuer son père lors d'une rencontre fortuite sur la route et épousera sa mère ensuite. La prophétie est finalement réalisée, malgré tous les choix entrepris pour l'éviter.

Or, si le déterminisme est vrai, nous avons encore de bonnes raisons de penser que nous ne partageons pas le destin d'Œdipe en ce que, contrairement à lui, nos choix et nos actions ont un réel pouvoir causal sur les futurs événements. Fatalement, Œdipe aurait toujours réalisé la prophétie, d'une façon ou d'une autre, et indépendamment de ses efforts. Or, selon le déterminisme, si Œdipe avait pu agir autrement, le futur aurait été modifié de telle manière qu'il aurait évité la prophétie, car si les causes sont modifiées, les effets le sont aussi. Au contraire, le fatalisme prétend que toutes les causes convergent vers un effet inéluctable, ce qui constitue un modèle métaphysique différent.

À la suite de cette distinction, nous pouvons commencer par écarter deux problèmes qui, par confusion, s'appliquent à tort au déterminisme, mais à raison au fatalisme. Une bonne partie des faux-problèmes partent effectivement de l'idée erronée selon laquelle le déterminisme, en nous empêchant de réellement faire ce que nous voulons, donne finalement lieu à la résignation, l'abandon et l'apathie.

Plus précisément, la première critique soutient que nos actions n'ont plus de sens, dans un monde où tout évènement est inéluctable. En effet, si le physicien d'Hawking croyait vraiment au déterminisme, il traverserait la route sans regarder, puisque sa mort est déjà déterminée par les lois de la nature en une date et un endroit précis. Il ne vaut pas la peine d'essayer d'empêcher ce qui est inéluctable : soit les lois de la nature prévoient qu'il sera percuté par une voiture, donc il n'y a pas lieu de regarder avant de traverser ; soit les lois de la nature prévoient qu'il ne sera pas percuté par une voiture, donc il n'y a pas non plus lieu de regarder avant de traverser. Il faut alors en conclure que le physicien déterministe est réduit à l'inaction. Toutefois, à y voir de plus près, on peut parvenir raisonnablement à cette conclusion uniquement si le physicien se trouve dans la situation d'Œdipe et non pas si le déterminisme est vrai. Dans un monde fataliste, le physicien n'aurait effectivement pas les moyens d'échapper à son destin, car tous ses choix ne peuvent mener qu'au même effet. Or, le déterminisme n'implique pas que nos actions n'aient pas de pouvoir causal sur la suite des évènements. Il est donc faux de penser que la prise de précautions du physicien est insensée. Au contraire, même si le déterminisme est vrai, les précautions permettent toujours d'éviter les dangers. Il reste vrai qu'un être humain qui ne fait pas attention en traversant la route a plus de chance de mourir. Le physicien regarde donc avant de traverser parce que cet acte préventif est *déterminant* et *efficace* pour éviter l'accident, le fait qu'il soit lui-même déterminé à agir ainsi n'empêchant pas l'efficacité de l'action elle-même. Dès lors, le problème disparaît à partir du moment où l'on saisit véritablement que l'efficacité causale reste permise dans un cadre déterministe. Les compatibilistes ont d'ailleurs été d'une grande utilité pour montrer que le déterminisme n'est pas incompatible avec toute forme de contrôle sur nos actions.

La seconde critique soutient que, même si nos actions demeurent peut-être efficaces, la délibération n'a plus de sens, quant à elle, et nous sommes donc contraints de ne plus délibérer, ce qui revient à ne plus agir. On peut comprendre cette nouvelle critique de deux manières. D'abord, en supposant que nos délibérations n'ont plus d'utilité si le déterminisme est vrai. Dans ce cas, nous pouvons répondre à cette objection en invoquant encore une fois la confusion entre fatalisme et déterminisme. En effet, nos délibérations demeurent aussi efficaces si le déterminisme est vrai puisqu'elles ont toujours un pouvoir causal.

La deuxième manière de formuler la critique, plus ambitieuse, concerne, quant à elle, l'irrationalité de la délibération déterministe. En effet, pour délibérer, nous devons croire que différents futurs sont ouverts et cette croyance est incompatible avec le déterminisme. La délibération est donc irrationnelle parce qu'elle implique la conjonction de deux croyances contradictoires.

Ici, il faut admettre que le déterminisme force une modification radicale dans nos croyances.

En effet, si le déterminisme est vrai, un seul futur est possible et nous ne pouvons plus persévéler dans l'idée qu'un ensemble de réalisations le sont. Il reste néanmoins possible d'argumenter que cette modification ne conduise pas nécessairement à un abandon radical de la délibération. En effet, croire qu'un seul futur est possible n'implique pas la connaissance de ce futur inéluctable, ni des choix que nous n'avons pas encore faits. En ce sens, les déterministes proposent de concevoir la délibération comme un processus épistémique, plutôt que métaphysique. En délibérant, les agents considèrent plusieurs possibilités épistémiques, alors qu'il n'existe qu'une seule possibilité métaphysique. Ainsi conçue, la délibération n'est pas nécessairement irrationnelle, puisque les croyances libertariennes sont substituées par des probabilités épistémiques. De la même manière qu'il n'est pas incohérent de croire 1) que la pièce de 5.- a 50% de chance de tomber sur pile ou face, et 2) que cette pièce ne peut tomber que sur l'une des deux faces, il n'est pas incohérent de délibérer en sachant qu'un seul futur est possible, parmi les options probables.

J'aimerais, à présent, identifier trois véritables problèmes qui s'appliquent au déterminisme dur, en me concentrant principalement sur le dernier. Un véritable problème consiste typiquement une pratique importante rendue incohérente par le scepticisme à propos de la responsabilité morale. Selon Peter Inwagen, le rejet de la responsabilité morale cause nécessairement ce type de problèmes, dans la mesure où nos attitudes courantes présupposent quotidiennement la responsabilité morale⁵⁸. Pour ce faire, je sépare tout d'abord deux types de problèmes : les problèmes moraux et les problèmes relatifs au sens de l'existence. Ceux-ci portent sur deux implications différentes du scepticisme à propos de la responsabilité morale. En dernier lieu, je montrerai que le problème des attitudes réactives est intimement lié au deux premiers.

3.1.1 Moralité

L'une des préoccupations premières, dès lors que l'on parvient à la conclusion que la responsabilité morale est impossible, est de se demander, en un sens large, comment la moralité peut exister dans ces conditions. En effet, sans responsabilité morale, nos actions ne sont plus susceptibles d'être blâmées ou louées. Il apparaît, dès lors, que nos pratiques morales sont considérablement mises en cause et doivent peut-être être raisonnablement suspendues. Il est certain que cette conclusion radicale est problématique pour le sceptique et mérite, à ce titre, un discours supplémentaire. Pour y voir plus clair, il est nécessaire d'identifier quelles sont les pratiques morales réellement menacées

⁵⁸ Van Inwagen, P. (1983). *Essai sur le libre arbitre*. (Trad. F. Nef). Vrin. (Traduction française de l'ouvrage "An Essay on Free Will")

par le scepticisme. Je vais, pour ce faire, passer en revue différentes formulations de cette difficulté générale.

Tout d'abord, on pourrait être tenté de juger que si la responsabilité morale n'existe pas, les concepts moraux de « bien » et « mal » n'ont plus lieu d'être. Néanmoins, les déterministes ont, à ce sujet, les ressources nécessaires pour défendre l'inverse. En effet, il serait étrange de soutenir qu'il est impossible de former des jugements axiologiques à propos d'une action donnée, sous prétexte qu'elle est déterminée. Imaginons un agent dont la tumeur au cerveau le force à adopter en permanence des comportements vicieux et cruels. Bien que nous reconnaissions volontiers qu'il serait injuste de le blâmer pour ses attitudes, il ne semble pas, pour autant, que ses actions deviennent moralement neutres. Ses actions vicieuses et cruelles ne remplissent pas les critères de l'action morale⁵⁹ et ce, indépendamment de la nature du processus causal qui a permis leur réalisation. Si tel est le cas, le déterminisme n'est pas incompatible avec l'usage des termes relatifs à l'évaluation morale.

Cela étant dit, bien que les actions puissent être morales ou immorales indépendamment du déterminisme, ce qui importe davantage, lorsque nous disons d'une action qu'elle est bonne ou mauvaise, n'est pas sa seule caractérisation morale, mais plutôt les récompenses et punitions qu'elle motive. En effet, il semble évident que nos pratiques morales ne consistent pas uniquement en des jugements froids, mais comprennent également d'autres attitudes qui impliquent le désir de punir ou récompenser l'agent qui le mérite. Une composante rétributionniste semble donc intimement liée au jugement moral. Or, si le déterminisme est vrai, personne n'est moralement responsable de ses actes et personne ne peut justement être puni ou récompensé. Cette difficulté représente des problèmes concrets au niveau du traitement des comportements criminels en particulier : si la responsabilité morale n'existe pas, nous n'avons plus de raison de punir les mauvais comportements. Ce point relève même d'une évidence philosophique selon Honderich :

"It is agreed on all hands that some factual belief about an action's having been free is required by us for holding an agent responsible and of course for blaming or punishing her. If I lose the belief, I cannot persist in the attitude or the behavior. »⁶⁰

Dès lors, l'une des difficultés du sceptique réside dans l'obligation d'abandonner l'idée selon laquelle nous punissons les criminels pour avoir librement mal agi. Mais est-ce à dire, néanmoins, que le rétributionnisme soit la seule manière de traiter convenablement les comportements criminels ?

⁵⁹ Peut-être est-ce néanmoins plus clair dans un cadre conséquentialiste que déontologiste. Les deux formulations de l'impératif catégorique, en termes de respect de la dignité et de principe de généralisation impliquent-elles l'autonomie ? Quant à l'éthique des vertus, je ne sais pas exactement comment me prononcer. Puisque l'on ne peut pas rattacher ce comportement au caractère de l'agent, j'ai l'impression que ce comportement serait exclu du cadre moral.

⁶⁰ Honderich, T. *Compatibilism, Incompatibilism, and the Smart Aleck*. Philosophy and Phenomenological Research 56 (1996), pp. 855–62, at pp. 856–9; cf. A Theory of Determinism.

En réalité, il est même permis de se demander si la punition, elle-même, réalise toujours une fin rétributioniste. En effet, lorsque l'on punit un enfant, par exemple, il est probable que l'acte n'ait pas nécessairement vocation à lui donner le mauvais traitement qu'il mérite, mais plutôt à favoriser un apprentissage et une éducation morale⁶¹. Les différents degrés de punition permettraient à l'enfant de saisir le degré de gravité de sa faute et le conduiraient à une meilleure conscience morale. De la même manière, il est possible de justifier la punition des criminels sur un motif similaire, en évacuant l'idée même de rétribution. Néanmoins, je crois que l'analogie entre la punition d'un criminel et celle d'un enfant doit être sujette à caution. En premier lieu, il est peu probable que tout criminel soit dans une situation d'ignorance enfantine qui nécessite un apprentissage moral. De plus, il n'est pas certain qu'une amélioration morale succède toujours à la punition. Toutefois, je crois qu'il est encore possible, pour le sceptique, de justifier traiter les comportements criminels, sans punition. En effet, sans même considérer la responsabilité morale, il est possible de justifier la privation de liberté des criminels, par souci de prévention ou de protection de la société, plutôt que par rétribution. Ainsi, les sceptiques pourraient défendre l'utilité d'une même mesure, tout en la justifiant différemment. La détention criminelle, ainsi conçue, se calquerait sur le modèle de la quarantaine : lorsqu'une personne est porteuse d'un virus dangereux, il semble, en effet, justifié de la tenir à l'écart pour préserver les autres. Le sceptique parviendrait ainsi à une approche conséquentialiste du droit :

"Suppose a person poses a danger to society by a sufficiently strong tendency to commit murder. Even if he is not in general a morally responsible agent, society would nevertheless seem to have as much right to detain him as it does to quarantine a carrier of a deadly communicable disease who is not responsible for being a carrier."⁶²

Contrairement au rétributionisme, cette forme de conséquentialisme est compatible avec le scepticisme à propos de la responsabilité morale, dans la mesure où elle ne présuppose pas nécessairement le libre arbitre et la responsabilité morale. En effet, l'usage de mesures sociales pour empêcher la réalisation de certaines tendances criminelles, n'implique pas nécessairement la croyance en une responsabilité à l'origine de ces tendances. Cette forme de conséquentialisme suppose seulement que la criminalité est un mal à prévenir socialement, au même titre qu'une maladie grave. En effet, le fait que les criminels ne méritent pas le blâme et la punition ne signifie pas pour autant qu'il serait juste de permettre leur liberté, alors qu'elle représente une menace. Bien que cette alternative puisse contrarier certaines de nos intuitions, il est peut-être préférable de l'adopter finalement, dès lors que l'on croit en l'impossibilité du libre arbitre.

Enfin, le problème moral le plus sérieux à mon sens, tient au rôle normatif que jouent nos énoncés moraux. En effet, on distingue généralement les faits des normes de la manière suivante : les faits

⁶¹ Pereboom, D. (2001). *Living without Free Will*. Cambridge University Press.

⁶² Ibid.

décrivent *ce qui est* et les normes prescrivent *ce qui doit être*. Or, si l'on comprend la morale comme un ensemble de prescriptions commandant ce qu'il *faut* et *ne faut pas* faire, il est probable que cette normativité n'ait plus vraiment de sens, dans un monde sans libre arbitre. En effet, il existe un principe kantien largement partagé selon lequel « *devoir* implique *pouvoir* ». Ce principe suggère que les agents doivent être capables de faire ce qui leur est moralement exigé. Or, si le déterminisme est vrai, personne ne *peut* agir autrement que ce qui est actuel et donc personne ne *doit* faire quoique ce soit. Je ne traiterai pas directement de cette question dans la suite de mes réflexions.

3.1.2 Le Sens de la Vie

Parallèlement aux problèmes moraux, il existe une inquiétude générale à propos du vide existentiel laissé par l'absence de responsabilité morale. En effet, si toutes nos décisions trouvent origine dans des facteurs externes à notre volonté, il semble finalement que nous ne contrôlons pas la direction que prennent nos existences, ce qui nous constraint à une position de spectateur de nos propres vies. À cause du déterminisme, nous sommes réduits à l'illusion de poursuivre les objectifs que nous croyons nous fixer librement. Aussi, les opposants au scepticisme prétendent, en définitive, que cette prise de conscience conduit à un état de renoncement relatif à l'existence elle-même et à un nihilisme vicieux. Toutefois, bien que le déterminisme entraîne sans doute une modification concrète au niveau de nos croyances à propos du monde, il est nécessaire de veiller à ne pas en surévaluer les conséquences, à échelle existentielle. Tout d'abord, il est possible de mettre en perspective ce préjudice, en soulignant les bénéfices que l'on peut également tirer, à titre personnel, de l'absence de libre arbitre. La perspective déterministe défendue par les stoïciens, par exemple, a pour vocation contraire de conduire à une attitude mesurée et une stabilité émotionnelle dont les avantages sont appréciés, plutôt que contestés. Par ailleurs, l'équanimité obtenue permet de se préserver des souffrances qui sont le signe d'une résistance ou lutte contre ce qui est inévitable. Selon les stoïciens, c'est au contraire par une conscience accrue du déterminisme que nos existences gagnent en valeur. Bien qu'il soit permis de douter des arguments stoïciens, je crois plus généralement qu'il est important de relativiser les effets généraux du déterminisme sur nos existences, en évitant de nourrir les poncifs habituels. Comme le souligne Daniel Dennett⁶³, le déterminisme pâtit d'une longue tradition de représentations inquiétantes et de métaphores connotées en philosophie. « Le marionnettiste », « la prison invisible » ou « le télépathe malveillant » sont effectivement des analogies qui contribuent à forger une certaine représentation qui n'est pas toujours fondée. Ces expressions forcent une conception selon laquelle le déterminisme serait le synonyme d'une manipulation externe ayant pour

⁶³ Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. MIT Press.

seul but de nous contraindre. Or, je ne crois pas que le déterminisme implique toujours une forme de coercition ou de manipulation au niveau de nos choix, ni qu'il soit intrinsèquement incompatible avec l'épanouissement et le sens dans l'existence. En effet, si le déterminisme impliquait nécessairement une coercition ou une manipulation, il apparaîtrait sans doute qu'une bonne partie de la valeur que nous accordons à nos vies serait perdue. Si nous apprenions, par exemple, que toutes nos décisions cruciales étaient le résultat d'une manipulation familiale orchestrée minutieusement depuis notre naissance, il apparaîtrait certainement que cette découverte aurait un impact très négatif sur la valeur et le sens de nos vies. Néanmoins, le déterminisme n'est pas systématiquement vécu comme une telle manipulation. Il semble, en effet, que les conditions génalogiques de la détermination, plutôt que la détermination elle-même, conditionne notre capacité à attribuer de la valeur à nos existences. Ainsi, dans certaines conditions, la détermination est jugée compatible avec l'épanouissement, alors qu'elle est inacceptable dans d'autres conditions. Dès lors, il reste possible que nos projets soient déterminés de telle façon que l'épanouissement est permis. Pereboom en propose un exemple qui me semble parlant :

"For example, suppose one discovered that one's career choice was determined by a deep interest, produced by the circumstances of one's upbringing, in finding solutions for some of the problems facing humanity. Or that the choice of one's spouse is explained by the salutary features of one's childhood environment, which one was determined to replicate because of the happiness it produced. These discoveries might well enhance the fulfillment provided by these projects."⁶⁴

Dans cet exemple, le déterminisme se défait relativement bien de la représentation inquiétante, développée plus haut. Si le type de déterminations auxquelles nous sommes le plus souvent soumis ressemblent à celles qui sont décrites dans l'exemple, alors nous avons de bonnes raisons de penser que le déterminisme n'est pas, en soi, incompatible avec l'épanouissement.

Toutefois, il est nécessaire de concéder que le déterminisme invalide une certaine conception de soi, qui est valorisée pour elle-même. L'idée selon laquelle nous sommes des agents libres, autonomes et capables de façonnner nos vies doit être effectivement revue, si le déterminisme est vrai. Le principal problème de cette révision est le passage abrupt d'une représentation selon laquelle l'être humain participe activement à sa vie, à une représentation selon laquelle il subit les événements. Je parlerai indirectement de ce problème dans la section prochaine.

3.1.3 Les Attitudes Réactives

J'en viens désormais au problème des attitudes réactives. Au chapitre I, j'ai défini les attitudes réactives comme un ensemble d'émotions qui attribuent de la responsabilité morale, en réaction aux

⁶⁴ Pereboom, D. (2001). *Living without Free Will*. Cambridge University Press.

bonnes et mauvaises volontés exprimées dans un contexte social. Or, si la responsabilité morale est impossible pour les raisons évoquées dans le chapitre précédent, nos attitudes réactives deviennent inadéquates, dans la mesure où elles représentent une propriété que l'objet ne possède pas. Par conséquent, elles doivent être abandonnées ou révisées, par souci de rationalité. Pourtant, cette conclusion est problématique pour les sceptiques à propos de la responsabilité morale, principalement pour deux raisons.

Tout d'abord, les sceptiques doivent admettre que l'abandon d'une pratique aussi quotidienne que l'attribution de responsabilité morale ne semble pas envisageable. Selon P.F. Strawson, il est même psychologiquement impossible de se débarrasser d'attitudes fondamentales et ordinaires telles que la colère, l'indignation, le remord ou la gratitude. En effet, ces attitudes structurent notre humanité en profondeur et il n'est pas certain que nous soyons assez sensibles aux raisons théoriques, en faveur du déterminisme, pour être capables d'une telle révision. Si tel est le cas, le sceptique doit expliquer pourquoi les attitudes réactives sont incontournables (*inescapable*), afin d'éviter l'incohérence.

Le second problème est le suivant : à supposer que nous en soyons capables, l'abandon des attitudes réactives représenterait un réel coût sur la vie humaine, étant donné le rôle essentiel qu'elles jouent dans nos existences. Ainsi, le scepticisme est pris en étau. En effet, soit nous sommes incapables d'abandonner les attitudes réactives et il revient au sceptique d'expliquer cette incapacité à vivre conformément à ses croyances ; soit nous sommes capables d'abandonner les attitudes réactives et le sceptique est dans l'obligation de reconnaître, au vu du caractère essentiel du rôle qu'elles jouent en termes moraux et existentiels, que cet abandon représente une trop grande perte. Comme j'ai tâché de l'évoquer précédemment, cette perte rejaillit principalement sur les deux autres types de problèmes que j'ai déjà identifiés.

Tout d'abord, les attitudes réactives réflexives telles que le remord, la culpabilité ou la honte occupent une fonction morale importante, parce qu'elles motivent un perfectionnement de soi. Grâce à ces émotions déplaisantes, nous prenons conscience de nos fautes et sommes disposés dans un second temps à les réparer. Sans attitudes réactives, il est probable que nous perdions un moyen de nous améliorer sur le plan moral :

“Shame, like guilt, is rightly viewed as an essential ingredient in what makes us moral agents. While guilt signals, and is justified by, our infringements of moral norms and typically motivates us to repair wrongs, shame signals and is justified by our failing to honor, even minimally, the demands consubstantial with the values to which we are attached, and it thus typically motivates us to reform ourselves.”⁶⁵

Par ailleurs, les attitudes réactives sont fondamentales et en ce qu'elles permettent une participation active aux relations et rapports interpersonnels. Selon cette vue, les attitudes réactives constituent le

⁶⁵ Deonna, J. Rodogno, R. et Teroni, F. *In Defense of Shame: The Faces of an Emotion* (2011; online edn, Oxford Academic, 19 Jan. 2012), <https://doi.org/10.1093/acprof:oso/9780199793532.001.0001>

coeur de notre sociabilité humaine. On peut, en effet, raisonnablement argumenter qu'en l'absence d'attitudes réactives personnelles telles que la gratitude, l'admiration ou le pardon, nous perdriions beaucoup de notre capacité à entretenir des relations centrales telles que l'amitié et l'amour. Il va sans dire qu'une telle perte affecterait le sens de nos existences. En effet, selon Strawson, il existe une différence fondamentale entre deux types d'attitudes relationnelles. Tout d'abord, *l'attitude subjective* est une attitude de sujet à sujet, qu'un adulte responsable adopte à l'encontre des autres adultes responsables. Cette attitude se caractérise par un certain nombre d'exigences interpersonnelles, qui se traduisent par la présence d'attitudes réactives. Au contraire, *l'attitude objective* est une attitude adoptée par les adultes responsables à l'encontre des enfants, des animaux ou des adultes atteints de maladie mentale sévère. Elle se caractérise, quant à elle, par une absence ou suspension des exigences interpersonnelles et des attitudes réactives. En effet, *l'attitude objective* semble être incompatible avec les exigences morales qui sont permises dans *l'attitude subjective*. Lorsque nous décrivons, par exemple, la mauvaise action d'une personne atteinte d'une déficience sévère par exemple, nous sommes effectivement plus enclins à adopter des termes biologiques plutôt que moraux. Aussi, nous réagissons le plus souvent à ces comportements sans attitudes réactives. Dès lors, si les attitudes réactives sont toujours inadéquates parce que le déterminisme est vrai, nous devons adopter une généralisation de *l'attitude objective* et concevoir toutes les actions humaines, non plus comme des actes voulus, mais plutôt comme des événements objectifs. Or, le principal problème de cette idée est que *l'attitude objective* semble conduire à des comportements froids et calculateurs qui sont incompatibles avec des relations humaines de qualité :

“Imagine for a moment what a world would be like in which we all regarded each other solely with the objective attitude. (...) The most gruesome difference between this world and ours would be reflected in our closest human relationships—in the relations between siblings, parents and children, and especially spouses and companions. We would still be able to form some sorts of association that could be described as relationships of friendship and love. One person could find another amusing or useful. One could notice that the presence of a certain person was, like the sound of a favorite song, particularly soothing or invigorating. We could choose friends as we now choose clothing or home furnishings or hobbies, according to whether they offer, to a sufficient degree, the proper combination of pleasure and practicality. (...) Nonetheless, I hope it is obvious why the words ‘friendship’ and ‘love’ applied to relationships in which admiration, respect, and gratitude have no part, might be said to take on a hollow ring. A world in which human relationships are restricted to those that can be formed and supported in the absence of the reactive attitudes is a world of human isolation so cold and dreary that any but the most cynical must shudder at the idea of it.”⁶⁶

Comme le suggère Susan Wolf, la substitution de *l'attitude subjective* par *l'attitude objective* représente une perte de valeur. Sans attitudes réactives, les relations personnelles semblent, en effet, perdre ce qu'elles avaient d'essentiel.

Au vu de ses larges implications, le problème des attitudes réactives devient un enjeu central pour les sceptiques à propos de la responsabilité morale. En premier lieu, comment expliquer notre incapacité

⁶⁶Wolf, S. (1981). *The Importance of Free Will*. Mind, 90(358), 386-405.

à abandonner les attitudes réactives ? Les sceptiques sont ici responsables d'une forme d'irrationalité qui consiste en l'incapacité de vivre en accord avec ses croyances.

Dans un deuxième temps, s'il est possible de suspendre les attitudes réactives, du moins pour les besoins de la réflexion, comment comprendre notre refus de le faire ? Le problème des attitudes réactives peut, ici, être compris comme un conflit entre deux types de rationalités : en un sens, il est rationnel de croire en la responsabilité morale au vu des avantages conséquents que cette croyance comporte à échelle humaine. Mais en un autre sens de « rationalité », il n'est pas rationnel de fonder une pratique sur une croyance que nous savons fausse, en dépit de ses avantages. Mon objectif, dans le reste de ce travail, sera d'essayer de trouver la meilleure réponse à ces questions.

3.2 Freedom and Resentment

Dans son influent *Freedom and Resentment*, P.F. Strawson se fixe pour objectif principal de réconcilier deux acteurs divergents du problème du libre arbitre, qu'il appelle les « optimistes » et les « pessimistes ». Bien que cette terminologie puisse sembler imprécise, il faut identifier, sous ces deux titres, l'opposition entre compatibilistes et incompatibilistes. Plus exactement, les « optimistes » de Strawson souscrivent à un genre spécifique de compatibilisme⁶⁷, qui prétend que nos pratiques morales telles que le blâme, la louange, la récompense et la punition se justifient par leur utilité sociale, qu'importe la vérité ou non du déterminisme. Ils sont donc dits « optimistes », au vu de leur tendance à relativiser les effets préjudiciables du déterminisme sur nos vies morales.

Les « pessimistes », à l'inverse, réfèrent aux incompatibilistes qui pensent que la vérité du déterminisme rend la responsabilité morale impossible et affecte donc toutes les pratiques qui la nécessitent. Ainsi, l'incompatibiliste est dit « pessimiste » parce qu'il considère que seul un libre arbitre libertarien est suffisant pour fonder la responsabilité morale, car l'utilité de la responsabilité morale n'est pas suffisante pour la responsabilité morale elle-même.

La position de Strawson, quant à elle, est comprise comme une voie de résolution, entre optimisme et pessimisme. D'un côté, Strawson considère, à l'instar des pessimistes, que la position optimiste est lacunaire et ne permet pas de justifier convenablement nos pratiques morales : analyser le blâme uniquement sur la base de son utilité sociale semble donner lieu, au mieux, à une forme de moralité insuffisante. En effet, en basant la responsabilité morale sur son utilité, il semble que les optimistes omettent quelque chose de vital, le concept de responsabilité morale n'apparaissant pas réductible à son utilité. Dans certaines situations, il est possible qu'il soit utile de tenir pour responsable un agent, alors qu'il ne l'est pas et, à l'inverse, il est possible qu'il ne soit pas utile de tenir pour responsable un agent qui l'est. Dès lors, en quoi l'utilité permettrait-elle de justifier le blâme ?

D'un autre côté, Strawson considère que les pessimistes, en se référant à une notion libertarienne obscure et trop exigeante conceptuellement (« the obscure and panicky metaphysics of libertarianism »), font appel à un moyen disproportionné de parer cette insuffisance.

⁶⁷ Selon cette forme de compatibilisme, ce n'est pas la responsabilité morale qui est compatible avec le déterminisme, mais plutôt l'ensemble de nos pratiques morales qui sont justifiées en vertu de leur utilité. Ce type de compatibilisme est une forme de « pragmatisme moral » ou de « conséquentialisme moral ».

Selon lui, la lacune peut être comblée plus simplement par nos attitudes réactives, car ce sont elles qui interviennent dans nos attributions de responsabilité morale :

“What is in question is the pessimist’s justified sense that to speak in terms of social utility alone is to leave out something vital in our conception of these practices. The vital thing can be restored by attending to that complicated web of attitudes and feelings which form an essential part of the moral life as we know it, and which are quite opposed to objectivity of attitude.”⁶⁸

Pour s’en convaincre, il faut néanmoins quitter l’analyse conceptuelle de la responsabilité morale, au profit d’une description psychologique de ce que nous faisons pratiquement lorsque nous tenons quelqu’un pour responsable. Strawson suggère, à ce titre, que cette nouvelle perspective nous mène à deux idées simples mais fondamentales : 1) la moralité ne peut être conçue hors de notre cadre psychologique et émotionnel, et 2) nous avons un attachement primitif aux attitudes réactives qui appartiennent à ce cadre. Ces deux constats donnent lieu, ensuite, aux deux thèses principalement défendues dans *Freedom and Resentment*. Dans un premier temps, Strawson milite pour une thèse psychologique qui soutient que nos attitudes réactives, en étant issues du cadre naturel de notre vie psychologique (« general framework of human life »), ne peuvent pas être remises en question par le déterminisme, lequel apparaît comme « hors-sujet », en regard de ce cadre naturel :

“The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them; and being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question.”⁶⁹

Dans un second temps, Strawson milite pour une thèse complémentaire, d’ordre normatif, qui soutient que, au vu de notre attachement primitif aux attitudes réactives et de leur rôle essentiel, ces attitudes ne doivent pas être remises en question par la croyance en le déterminisme :

“And I shall reply, second, that if we could imagine what we cannot have, viz. a choice in this matter, then we could choose rationally only in the light of an assessment of the gains and losses to human life, its enrichment or impoverishment; and the truth or falsity of a general thesis of determinism would not bear on the rationality of this choice.”⁷⁰

Ainsi, la véritable contribution de Strawson repose sur une idée qui n’avait pas été suggérée jusqu’alors : on peut résoudre le problème des attitudes réactives, par les attitudes réactives elles-mêmes. Néanmoins, le succès de cette solution nécessite, en premier lieu, que nous quittions une approche commune à toutes les thèses traditionnelles de la responsabilité morale qui, malgré les désaccords, partagent toutes la même erreur.

⁶⁸ Strawson, P. F. (1962). *Freedom and Resentment*. Proceedings of the British Academy, 48, 187-211.

⁶⁹ Ibid.

⁷⁰ Ibid.

En effet, compatibilistes et incompatibilistes se trompent en cherchant rigoureusement les conditions nécessaires et suffisantes pour la *responsabilité morale* et oublient de considérer l'aspect psychologique de *l'attribution de responsabilité morale*, dans lequel nos attitudes réactives sont centrales. En localisant cette erreur, Strawson pense finalement réconcilier les acteurs principaux du débat du libre arbitre.

3.2.1 Thèse Naturaliste

Selon Strawson, l'erreur fondamentale des thèses traditionnelles repose sur l'analyse théorique qui est fournie d'emblée dans la recherche des conditions suffisantes au concept, alors que la responsabilité morale est une notion essentiellement pratique, qu'il s'agit de décrire et délimiter, plutôt qu'analyser. En y regardant de plus près, c'est en effet par le biais de nos attitudes réactives que nous attribuons de la responsabilité morale, car ce sont ces mêmes émotions qui interviennent pratiquement dans l'usage que nous en faisons. Aussi, nous concevons les agents comme responsables, non pas sur la base d'une analyse métaphysique préalable, mais plutôt grâce à la structure de nos engagements naturels dirigés vers les autres participants de notre communauté morale et nos réactions primitives aux intentions manifestées par ces-derniers. Dès lors, fonder la responsabilité morale sur des conditions métaphysiques robustes entraîne, de façon problématique, la sur-intellectualisation d'une notion essentiellement ordinaire. En effet, dans nos pratiques, les attributions de responsabilité morale reposent, plus modestement, sur des attitudes primitives auto-justificatrices, au même titre que certaines formes de sympathie, d'animosité et d'amitié sont des attitudes qui ne semblent pas nécessiter de justification en dehors de leur cadre.

Les attitudes réactives sont l'expression naturelle d'une préoccupation fondamentale dirigée vers l'évaluation des intentions d'autrui et n'ont pas besoin d'être justifiées autrement :

"It matters to us (...) whether the actions of other people (...) reflect attitudes toward us of good will, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other"⁷¹

À ce titre, chercher à justifier l'attribution de la responsabilité morale, en dehors des conditions propre à son usage est une erreur philosophique et un excès de rationalisation. Au quotidien, si nous faisons tous usage de cette notion correctement, ce n'est pas sur la base d'un principe théorique, tel que la possibilité alternative ou la source agentive, mais simplement grâce à la structure interne de nos

⁷¹ Strawson, P. F. (1962). *Freedom and Resentment*. Proceedings of the British Academy, 48, 187-211.

attitudes. Les attitudes réactives sont donc suffisantes pour concevoir la responsabilité morale et la lacune est comblée.

Selon Strawson, cette structure autonome est capable, par ailleurs, d'opérer les distinctions qui nous importe, dans les attributions de responsabilité morale. Il existe, donc, des façons appropriées de blâmer et louer les personnes même dans un cadre déterministe. Bien qu'il nous arrive de suspendre ou modifier nos attitudes réactives, sous certaines conditions propres à nos pratiques, le déterminisme, en tant que thèse abstraite, n'est jamais véritablement pertinent pour savoir si telle attitude est appropriée ou non.

Dès lors, Strawson se propose de décrire nos usages, en identifiant les considérations qui entrent typiquement en jeu dans la suspension ou la modification de nos émotions. En premier lieu, il apparaît que nous suspendons ou modifions nos attitudes réactives, seulement lorsque l'agent initialement visé par nos réactions bénéficie de bonnes défenses («pleas »⁷²). Il existe essentiellement deux types de défenses qui permettent une révision de nos attitude réactives : les excuses et les exemptions. Les excuses produisent une suspension ou modification des attitudes réactives en montrant que, même si une certaine conduite est mauvaise, elle ne manifeste cependant aucune mauvaise volonté. En effet, les excuses révèlent que la conduite en question ne manifeste pas, contrairement aux apparences, de la mauvaise volonté. On formule généralement les excuses à l'aide des phrases suivantes : « Je n'avais pas l'intention de », « je ne pensais pas que », « Je ne savais pas que », etc...⁷³ Imaginons, par exemple, une personne qui en bouscule une autre, sans aucune mauvaise intention. Dans ce cas, il y a de fortes chances que la colère ou l'indignation soit effectivement jugée inappropriée, après que l'excuse est formulée et la mauvaise volonté écartée.

Les exemptions, quant à elles, produisent une suspension ou modification des attitudes réactives en montrant que, même si une certaine conduite est mauvaise et exprime une mauvaise volonté, l'agent en question est incapable d'entretenir des relations interpersonnelles normales et de participer à la communauté morale. Les exemptions sont typiquement invoquées pour exclure certaines catégories d'êtres qui, par leur invalidité psychologique ou leur immaturité, sont perçues comme inaptes à la moralité. On formule généralement les exemptions à l'aide des phrases suivantes : « Ce n'est qu'un enfant », « Il est malade », « Il n'est plus lui-même », etc...⁷⁴ Imaginons, par exemple, une personne victime d'une forme sévère de schizophrénie, mettant le feu à des habitations. Dans ce cas, il y a également de fortes chances que notre colère ou indignation soit suspendue, après que l'exemption

⁷² McKenna, M. S. (1998). *The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism*. The Journal of Ethics, 2(2), 123–142. Il est difficile de traduire le terme « plea » autrement que par défense, étant donné que le terme juridique « plaidoyer » paraît inapproprié dans ce contexte.

⁷³ Ibid.

⁷⁴ Ibid.

est formulée. En effet, les exemptions conduisent à adopter une attitude objective envers leur cible : nous voyons en elle un défaut ou une incapacité à s'engager normalement dans les relations interpersonnelles. Selon Strawson, ces deux types de défenses sont les seules raisons capables de motiver la suspension ou la modification de nos attitudes réactives. Dès lors, même si le déterminisme est vrai, celui-ci ne constitue jamais une excuse, ni une exemption, aux mauvaises conduites que nous adoptons. Premièrement, le déterminisme n'est pas une excuse, parce qu'il n'empêche pas nos actions d'exprimer de mauvaises intentions. Deuxièmement, le déterminisme ne constitue pas une exemption, parce qu'il n'implique pas le fait que tout le monde soit psychologiquement inapte. Ainsi, le déterminisme ne fournit pas de motif universel pour saper le bienfondé de nos attributions ordinaires de moralité.

En résumé, on peut clarifier la stratégie naturaliste de Strawson en quelques points clefs.

Tout d'abord, la qualité des intentions et volontés exprimées par les autres nous importe beaucoup. Ce souci donne lieu à des exigences qui définissent si un comportement est approprié ou non. À ce titre, les attitudes réactives se conçoivent comme des réactions à la satisfaction ou non de ces exigences.

Ensuite, Strawson suggère que ces réactions sont sujettes à des modifications, et parfois même à des suspensions. Il existe deux raisons de réviser nos attitudes réactives : les excuses et les exemptions. Enfin, ces deux types de défenses révèlent des différences dans les intentions et capacités des différents agents. Les attitudes réactives appropriées sont celles qui traquent ces différences. Or, ces différences ne sont pas annulées par la vérité du déterminisme. En effet, il existe toujours une différence de capacité entre un enfant et un adulte, indiquée par notre pratique de l'exemption. Dès lors, si la vérité du déterminisme n'implique pas une incapacité générale à s'engager dans des rapports interpersonnels, ni le fait que tout le monde ait de bonnes intentions, alors cette thèse n'est pas pertinente.

3.2.2 Naturalisme et Responsabilité Morale

Une autre question qu'il est possible de se poser à propos de la thèse naturaliste défendue dans *Freedom and resentment* est la suivante : Strawson offre-t-il une définition de la responsabilité morale ? Plus précisément, Strawson explique-t-il ce que c'est, pour un agent, d'être moralement responsable ? Et, si oui, quelles sont les conditions à remplir pour être moralement responsable ? À proprement parler, je crois qu'il n'existe, dans l'article, aucune évidence littérale allant dans le sens d'une telle définition. Dès l'entame de son article, Strawson se fixe un but différent : réconcilier les optimistes et les pessimistes. Toutefois, il est vrai qu'une connexion est clairement établie entre les attitudes réactives et la responsabilité morale. Certains commentateurs influents ont alors

interprété Strawson comme attribuant une priorité métaphysique à l'attribution de responsabilité morale, sur la responsabilité morale elle-même⁷⁵ Watson, par exemple, propose une interprétation selon laquelle Strawson renverrait effectivement la logique de la thèse traditionnelle de la responsabilité morale, pour proposer ensuite une nouvelle définition de la responsabilité morale (voir la section 1.2.).

Selon cette nouvelle définition, être responsable s'explique par rapport aux réactions émotionnelles exprimées par les autres. Ce n'est donc pas la responsabilité morale qui expliquerait l'attribution de responsabilité morale, mais plutôt l'inverse :

"In Strawson's view, there is no such independent notion of responsibility that explains the propriety of the reactive attitudes. The explanatory priority is the other way around: It is not that we hold people responsible because they are responsible; rather, the idea (our idea) that we are responsible is to be understood by the practice, which itself is not a matter of holding some propositions to be true but of expressing our concerns and demands about our treatment of one another."⁷⁶

Cette interprétation de la thèse strawsonienne peut ensuite donner lieu à différentes analyses de la responsabilité morale qui s'appuient sur cette priorité métaphysique et définissent la responsabilité morale à partir des attitudes réactives :

- Constructivisme : X est responsable de P, si X est tenu responsable de P.
- Dispositionalisme : X est responsable de P, si X est *disposé* (dans certaines circonstances) à être tenu responsable de P.
- « Fittingness account »⁷⁷ : X est responsable de P, si X est tenu responsable de P, *de façon appropriée*.

Si l'interprétation est correcte, alors il est effectivement possible de déduire ce type de définitions du discours de Strawson. Bien que le philosophe ne le formule pas clairement ainsi, certains commentateurs pensent néanmoins que ces définitions sont d'inspiration strawsonienne.

Pourtant, selon d'autres commentateurs, en particulier Maria Alvarez⁷⁸, Strawson, en écrivant *Freedom and Resentment*, n'a jamais voulu proposer une thèse de la responsabilité morale.

⁷⁵ Voir: Watson G. (2014). *Peter Strawson on Responsibility and Sociality*. Oxford Studies in Agency and Responsibility, 2, pp. 15–32. / McKenna, M. (2012). *Conversation and Responsibility*. New York: Oxford University Press. / Shoemaker D. (2017). *Response-Dependent Responsibility; or, a Funny Thing Happened on the Way to Blame*. Philosophical Review, 126(4), pp. 481–527.

⁷⁶ Watson, G. (1987). *Responsibility and the Limits of Evil: Variations on a Strawsonian Theme*. In Schoeman (Ed.), *Responsibility, Character, and the Emotions*. Cambridge University Press.

⁷⁷ McCormick, K. (2022). *The Problem of Blame: Making Sense of Moral Anger*. Cambridge: Cambridge University Press. doi:10.1017/9781108907071

⁷⁸ Alvarez, M. (2021). P. F. Strawson, Moral Theories and ‘The Problem of Blame’: ‘Freedom and Resentment’ Revisited. Aristotelian Society Supplementary Volume 95 (1):183–203.

Selon cette dernière, la connexion entre attitudes réactives et responsabilité morale, effectivement établie par Strawson, n'implique cependant pas la priorité explicative des attitudes réactives sur la responsabilité morale. Cette connexion est plutôt un moyen argumentatif de montrer que le concept de responsabilité morale n'est pas mis en péril par le déterminisme.

Dans la mesure où cette question herméneutique reste ouverte à mon sens, je ne prendrai pas position. Par ailleurs, la suite de mes réflexions ne nécessite pas de réponse fixe en la matière.

3.2.3 Thèse Normative

À la suite de la thèse naturaliste, il est possible d'attribuer à Strawson, en second lieu, une thèse normative correspondante. Non seulement le déterminisme n'a aucun impact sur nos attitudes réactives, mais encore il *ne doit pas* avoir un tel impact. Cette deuxième thèse sert à prévenir une critique potentielle qui objecterait que, bien que la thèse naturaliste décrive convenablement nos attributions de responsabilité morale, elle ne justifie pas pour autant de telles pratiques. En effet, une pratique peut *être* d'une certaine manière, sans *devoir être* telle qu'elle est. Selon Strawson, cette objection est néanmoins infondée, dans la mesure où il serait insensé de soutenir que nos attitudes réactives doivent être autres que ce qu'elles sont. La thèse naturaliste prétend précisément que le cadre naturel qui gouverne nos attitudes réactives est incontournable (sauf défenses) et ne peut pas être infléchi. En un sens, nous n'avons pas à justifier ce que nous sommes.

On peut néanmoins interpréter l'article de Strawson comme concédant un autre argument qui fait appel à la rationalité :

"It is a question about what it would be rational to do if determinism were true, a question about the rational justification of ordinary inter-personal attitudes in general. Such a question could seem real only to one who has utterly failed to grasp the purport of the preceding answer (...). the commitment [to ordinary inter-personal relationships] is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general framework."⁷⁹

En supposant qu'il serait possible que le déterminisme entraîne la suspension de nos attitudes réactives, cela impliquerait-il, pour autant, qu'il serait rationnel d'adopter une telle suspension ?

Nos attitudes réactives se verrraient alors substituées par une attitude objective similaire à celle que nous adoptons envers les enfants :

"To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided (...) The objective attitude may be emotionally toned in many ways: it may include repulsion or fear, it may include pity or love, though not all kinds of love. But it cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in interpersonal human relationships; it

⁷⁹ Strawson, P. F. (1962). Freedom and Resentment. Proceedings of the British Academy, 48, 187-211.

cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other.”⁸⁰

Or, il va sans dire qu'une telle substitution, au-delà du fait qu'elle est impossible partout et en permanence, ne serait en aucun cas souhaitable. Si nous avions effectivement la capacité d'adopter une attitude objective généralisée, cela n'impliquerait cependant pas qu'il serait rationnel d'adopter une perspective si radicale qu'elle entraînerait la fin des relations humaines auxquelles nous tenons. En effet, peut-être Strawson veut-il arguer également qu'une théorie aussi spéculative que le déterminisme n'est pas le bon type de raison pour abandonner les attitudes réactives. Le point, ici, est qu'une description essentiellement théorique du monde n'est pas à même de fournir les raisons pratiques suffisantes pour abandonner nos attributions de responsabilité morale. Le déterminisme est une doctrine théorique qui possède des implications théoriques, mais ne mène pas à des conséquences pratiques. Dès lors, si nous devions choisir d'abandonner les attitudes réactives, seules des raisons pratiques entreraient en considération. Mais, si l'on raisonne pratiquement, en considérant ce que nous avons à gagner ou à perdre, la généralisation de l'attitude objective représenterait une trop grande perte sur nos vies. Par conséquent, il serait plus rationnel de conserver les pratiques que nous jugeons essentielles :

“It would not necessarily be rational to choose to be more purely rational than we are.”⁸¹

3.2.4 Comment Strawson répond-il aux problèmes des sceptiques ?

Plus haut dans ce chapitre, j'ai formulé le problème des attitudes réactives de deux manières. Premièrement, s'il nous semble impossible, à première vue, d'adopter l'attitude objective en permanence, existe-t-il un moyen d'expliquer en quoi le déterminisme n'est pas pertinent pour l'adoption de l'attitude objective ? Ensuite, existe-t-il un moyen de résoudre la tension entre les deux formes de rationalités que les sceptiques ont identifiées ? Je vais désormais rappeler dans quelle mesure Strawson fournit une réponse directe à mes questions. En réalité, il est intéressant de se demander d'abord si le philosophe formule une ou deux réponses. Cela dépend essentiellement de la façon que nous avons d'envisager la relation entre les deux thèses étudiées, laquelle ne me paraît pas entièrement explicitée. La première réponse qu'il faut attribuer à Strawson peut être déduite de sa thèse naturaliste. Dans la mesure où le déterminisme ne peut pas affecter le cadre naturel relatif à l'attribution de responsabilité morale, le naturalisme répond aux deux questions. En effet, si l'on comprend bien, premièrement, la définition naturaliste suggérée par Strawson, il est naturellement

⁸⁰ Ibid.

⁸¹ Ibid. (note de bas de page)

impossible, pour un être humain, d'abandonner ses attitudes réactives à cause du déterminisme, car seules les exemptions et les excuses fournissent un réel motif. C'est pour cette raison que le déterminisme n'est pas pertinent. Deuxièmement, si l'on comprend bien la définition naturaliste de Strawson, il n'y a pas lieu de résoudre une tension qui se dissout d'elle-même à partir du moment où l'on saisit que la responsabilité morale est incontournable.

La deuxième réponse est déduite de la thèse normative qui, bien que secondaire, peut probablement être attribuée à Strawson : par raisonnement pratique, le déterminisme ne doit pas affecter nos pratiques morales et nous devons maintenir notre croyance en la responsabilité morale. Cette thèse secondaire offre ainsi une réponse à la deuxième formulation du problème des attitudes réactives : puisque la croyance théorique au déterminisme n'est pas pertinente dans nos raisonnements pratiques, seules les raisons pertinentes au maintien du statu quo entrent véritablement en considération. La tension est alors résolue. Néanmoins, il n'est pas facile de savoir si cet argument nécessite, en premier lieu, la validité de la thèse naturaliste de la responsabilité morale. La question est, ici, celle de la dépendance de la thèse normative vis-à-vis de la thèse naturaliste. En effet, peut-on établir d'emblée qu'il est irrationnel de modifier nos attitudes réactives sur la base de notre croyance au déterminisme (thèse normative), sans établir d'abord que cette croyance n'est pas pertinente (thèse naturaliste) ? Si ce n'est pas le cas, il apparaîtra alors que la thèse normative n'apporte aucune plus-value au niveau de la résolution du problème. Au contraire, si la thèse normative ne dépend pas de la thèse naturaliste, alors nous devons considérer que Strawson propose deux réponses au problème des attitudes réactives. Il sera question de cette distinction plus loin.

3.3 Critiques de la thèse naturaliste

La thèse naturaliste a reçu de nombreuses critiques qui ont partiellement remis en question sa popularité philosophique. J'en présente trois qui me semblent intéressantes. La première concerne les définitions de la responsabilité morale d'inspiration naturaliste que j'ai présentées dans la section **(3.2.2.)**, alors que les deux dernières portent sur le naturalisme en général.

Selon Fischer et Ravizza⁸², tout d'abord, le blâme moral ne peut pas être conçu uniquement comme le fait de recevoir la désapprobation des participants d'une communauté morale donnée. En effet, une telle théorie ne semble pas faire la différence entre *être tenu responsable de X* et *être effectivement responsable de X* :

« After all, once the actual application of the reactive attitudes is taken to be constitutive of moral responsibility, one wonders what should be said about situation in which communities hold people responsible who intuitively are not. Does the mere fact that certain attitudes are taken toward an agent establish that he is an appropriate candidate for this treatment? ”

Pour montrer cette distinction fondamentale, Fischer et Ravizza imaginent une société dans laquelle les personnes présentant des retards mentaux sont systématiquement exposés à la colère et au blâme des autres membres de la communauté. Ils sont même punis pour cette seule raison. Dans ce cas fictif, les attitudes réactives qui émanent de la communauté morale suffisent-elles pour rendre ces personnes responsables de leurs troubles et pour justifier leur punition ?

Inversement, imaginons une société dans laquelle certains traits physiques seraient perçus comme un critère d'exclusion de la communauté morale et justifieraient une attitude objective à leur égard, c'est-à-dire une attitude similaire à celle que nous adoptons auprès des jeunes enfants et qui nous empêche, par exemple, de nous mettre en colère contre eux. Une telle absence d'attitudes réactives contre le groupe concerné suffirait-elle néanmoins à invalider sa responsabilité morale, en raison de caractéristiques pour le moins arbitraires ?

Dans les deux cas, Fischer et Ravizza mettent en lumière la différence manifeste que nous faisons intuitivement entre *être tenu responsable* et *être réellement responsable*. En effet, il est possible *d'être tenu responsable, sans être réellement responsable* (premier cas) et inversement, *d'être réellement responsable sans être tenu responsable* (deuxième cas). Or, cette lacune théorique semble poser un

⁸²Fischer, J.M et Ravizza M. (eds.) 1993: *Perspectives on Moral Responsibility*. Ithaca, NY: Cornell University Press.

véritable problème sur le plan de la justice. En effet, est-il juste de punir les personnes présentant des retards mentaux, sous l'unique prétexte qu'elles provoquent la désapprobation des personnes « normales » ? Pour être puni justement, il semble que l'occurrence d'attitudes réactives ne soit pas suffisante et qu'il soit nécessaire de faire appel à une notion de mérite.

Cependant, les défenseurs de la thèse naturaliste de la responsabilité morale peuvent précisément répondre que cette objection s'applique au constructivisme et au dispositionalisme seulement, et non pas à l'analyse en termes de « fittingness » (voir **3.2.2.**).

En réalité, la majorité des strawsonniens défendent cette troisième version de la responsabilité morale, qui a pour avantage de faire place au mérite. En termes de « fittingness », seul un agent tenu responsable, *de façon appropriée*, est moralement responsable. Donc, pour qu'un agent soit moralement responsable, il ne suffit pas seulement qu'il fasse l'objet d'attitudes réactives, mais encore faut-il que celles-ci soient appropriées.

Par ailleurs, rappelons que Strawson remarque que les conditions de l'attribution de responsabilité morale sont fixées par notre usage de l'excuse et de l'exemption, lequel révèle des différences de capacités entre les agents. Or, les personnes victimes d'atteinte mentale sévère sont généralement sujettes à l'exemption. Dès lors, une société qui punirait certains handicaps, en ne faisant pas les distinctions pertinentes, permettrait des attributions de responsabilité morale inappropriées. Par conséquent, le problème est vraisemblablement résolu.

Néanmoins, je crois quand même que Fischer et Ravizza attirent notre attention sur un point plus général. En effet, rappelons d'abord que la norme qui définit si une attitude est appropriée ou non, dépend nécessairement des pratiques d'une communauté morale donnée (Framework). Dès lors, dans la perspective de Strawson, il est, *a priori*, impossible qu'une communauté commette des erreurs systématiques en traitant ses participants comme les cibles légitimes des attitudes réactives, alors qu'ils ne le sont pas, et vice versa. Or, si une société commettait de telles erreurs (ce qui ne semble pas impossible en soi), il serait alors impossible de critiquer ses pratiques. Etant donné la dépendance de la norme vis-à-vis de la communauté morale, il n'est pas permis de critiquer la structure entière, compte tenu de notre incapacité à adopter un point de vue objectif, hors de cette structure. Le problème est donc plus largement, celui du décalage entre l'impossibilité théorique de la critique sociale (issue du modèle strawsonien) d'une part, et l'existence attestée de la critique sociale d'autre part. Je ne suis pas certain que Strawson ait les ressources théoriques pour combler cette lacune ?

Cela étant dit, si comme le prétend Maria Alvarez⁸³, on ne peut pas attribuer à Strawson la définition selon laquelle la responsabilité morale est une attribution de responsabilité morale appropriée, alors l'objection que je viens de présenter n'est pas pertinente. Dès lors, je vais désormais tâcher de présenter des critiques qui s'attaquent plus directement au naturalisme.

En effet, certains philosophes rejettent en bloc la conclusion selon laquelle le déterminisme n'a aucun effet sur nos attitudes réactives. Tout d'abord, selon Saul Smilansky⁸⁴, il n'est pas vrai que toutes les attitudes réactives sont insensibles à la vérité du déterminisme. En effet, la thèse naturaliste de Strawson ne parviendrait pas à rendre compte adéquatement de certaines attitudes réactives complexes comme le « respect de soi ». Si l'exemple de la colère, qui se manifeste comme une émotion particulièrement spontanée et peu exigeante sur le plan cognitif⁸⁵, semble abonder dans le sens de la non-pertinence du déterminisme dans nos pratiques, il est néanmoins possible que le « respect de soi », qui se présente comme une attitude mesurée et requérant un système de croyances cognitivement complexes, soit quant à lui dépendant de certaines croyances théoriques. En d'autres termes, les croyances métaphysiques à propos du libre arbitre ne semblent pas occuper la même place pour ces deux attitudes réactives. Elles pourraient même être insignifiante pour la colère, mais pertinentes pour le « respect de soi ». En effet, étant donné le caractère spontané et cognitivement peu exigeant de la colère, il est peu probable qu'une croyance complexe comme celle du déterminisme puisse motiver sa suspension. À l'inverse, étant donné le caractère mesuré et cognitivement complexe du « respect de soi », il semble tout aussi improbable qu'une croyance au déterminisme, en considération de ses implications sur notre contrôle, la moralité et le sens de l'existence, laisserait l'émotion intacte. Grâce à ce contre-exemple, Smilansky remet donc en question la thèse selon laquelle la croyance au déterminisme n'est pas pertinente dans toutes nos attitudes réactives.

On trouve une objection similaire chez Gary Watson⁸⁶, lequel soutient, de façon plus ambitieuse encore, que toutes les attitudes réactives, à commencer par la colère, sont susceptibles d'être atténuées, modifiées, voire suspendues par une croyance au déterminisme. Pour ce faire, Watson note d'abord, en prenant le cas d'un célèbre meurtrier américain, Robert Harris, coupable d'avoir tué deux jeunes hommes de sang-froid et voler leur voiture, que nous avons effectivement une réaction d'indignation et de colère devant ces actes atroces.

⁸³ Alvarez, M. (2021). *P. F. Strawson, Moral Theories and 'The Problem of Blame': 'Freedom and Resentment' Revisited*. Aristotelian Society Supplementary Volume 95 (1):183-203.

⁸⁴ Smilansky S. *Free Will: From Nature to Illusion*. Proceedings of the Aristotelian Society, 101 (2001), 71–95.

⁸⁵ L'émotion est généralement déclenchée très soudainement.

⁸⁶ Watson, G. *Responsibility and the Limits of Evil: Variations on a Strawsonian Theme*. Schoeman (ed.), Responsibility, Character, and the Emotions (1987).

Ces attitudes réactives sont d'autant plus vives que le profil vicieux, froid et cruel du meurtrier est indiscutable. Après le meurtre, Robert Harris serait tranquillement rentré chez lui, à bord du véhicule volé, et aurait fini les cheeseburgers entamés par ses victimes, en fanfaronnant auprès de son frère et en se moquant du désespoir des deux garçons au moment de la mise à mort. En étudiant ce cas précis, il ne fait aucun doute que Robert Harris incarne l'archétype même d'un agent mal intentionné, aux volontés vicieuses et capable de susciter les attitudes réactives les plus tranchées. Cependant, Watson montre, ensuite, qu'une connaissance approfondie des déterminations socio-culturelles qui sont à l'origine de son acte criminel peuvent avoir un effet limitant sur nos attitudes réactives. En effet, en prenant connaissance de l'enfance monstrueuse du meurtrier, de l'alcoolisme de ses parents, de son abandon maternel, des maltraitances et humiliations omniprésentes, ainsi que des viols successifs qu'il a subis, nous atténurons, modifions voire suspendons nos réactions d'indignation et de colère. Le déterminisme est donc, au moins en partie, capable d'une certaine influence sur nos attitudes réactives. Pourtant, il est clair que la détermination ne diminue en rien la brutalité et la cruauté de ses actes. Pourquoi, dès lors, sommes-nous sensibles à ce genre de récit ? Ici, la réponse la plus probable est que nos réactions sont modifiées sous l'effet de l'explication qui est fournie. Selon Watson, nous concevons en Harris une incapacité, similaire à celle qui perçue dans les cas d'exemptions. En effet, Robert Harris nous paraît inapte à certaines capacités fondamentales telle que l'amitié, la sympathie et les considérations morales. Sa haine et son insensibilité sont donc conçues comme une incapacité qui l'exclut de la communauté morale et c'est pour cette raison que nos attitudes réactives sont modifiées. Néanmoins, cette conclusion est problématique pour la thèse de Strawson. En effet, est-ce à dire que l'acte d'Harris est si parfaitement méchant, qu'il n'est pas plus responsable ? Si la cruauté la plus pure est conçue ici comme une incapacité morale, alors tout acte méchant est susceptible de motiver une exemption :

“However, to exempt Harris on these grounds is problematic. For then everyone who is evil in Harris's way will be exempt, independently of facts about their background.”⁸⁷

Enfin, une dernière faiblesse de la théorie naturaliste de la responsabilité morale tient précisément à l'*appel à la nature* auquel elle se réfère. S'il est vrai que nous sommes naturellement enclins à réagir émotionnellement aux comportements des autres participants de la communauté morale et si nous avons une propension primitive à les tenir pour responsables de leurs actes, cela ne signifie pas pour autant que nous *devons* agir de la sorte. De plus, il se pourrait que nous ayons une propension tout aussi naturelle à juger que le déterminisme et ses implications concrètes au niveau de notre contrôle jouent un rôle limitant dans nos attributions de responsabilité. En effet, il est difficile de se fier

⁸⁷ Ibid.

essentiellement à la nature de nos engagements primitifs, si l'on reconnaît que ceux-ci peuvent être plus complexes qu'il n'y paraît :

"One might again object, that while we have a deep and perhaps inderacinable commitment to the reactive attitudes and practices, it is also in our nature to take determinism to pose a serious problem for our notions of responsibility and freedom. (...) Our commitments are complex, and conflict. Although our thoughts about determinism appear in actual fact quite impotent to disturb our natural and unconsidered reactive attitudes and feelings (this reveals one commitment), it also seems very difficult for us not to acknowledge that the truth of determinism or of non-self-determinability brings the propriety of the reactive attitudes seriously into doubt (this reveals the other commitment)."⁸⁸

Dès lors, si nos engagements naturels sont de deux natures distinctes en matière de responsabilité, il serait étrange de défendre une thèse naturaliste qui ne prenne en compte que l'une des deux natures. Par conséquent, au vu des sérieuses difficultés rencontrées par la thèse naturaliste de la responsabilité morale, je considère que le problème que j'ai identifié n'est pas encore résolu par Strawson.

⁸⁸ Strawson, G. (1993). *Freedom and Resentment*. In John Martin Fischer & Mark Ravizza (eds.), Perspectives on Moral Responsibility. Cornell University Press. pp. 67-100.

3.4 Critiques de la Thèse Normative

Qu'en est-il de la thèse normative ? constitue-t-elle une réponse satisfaisante au problème des attitudes réactives ? Comme je l'ai proposé dans la section (3.2.4.), nous pouvons distinguer deux interprétations de la thèse. Tout d'abord, si la thèse normative dépend de la thèse naturaliste (c'est à mon sens l'idée initiale de Strawson), alors elle se heurte aux mêmes problèmes. Néanmoins, si on considère, au contraire, que les deux thèses sont indépendantes, il me faut encore étudier une autre solution.

Rappelons la thèse normative : nous avons un attachement certain aux attitudes réactives et nous reconnaissions le grand intérêt que nous avons à les maintenir. Ainsi, par un raisonnement pratique, selon lequel un acte est jugé rationnel s'il est un moyen de parvenir à une fin⁸⁹, nous *devons* persévérer dans nos attitudes réactives, parce qu'elles sont un moyen de parvenir à des fins humainement centrales comme l'amitié ou l'amour. La prémissé de cet argument est que le déterminisme n'est pas le bon type de raison pour abandonner les attitudes réactives. On a coutume, en effet, de distinguer deux types de raisonnements : les raisonnements pratiques et théoriques. Le premier type de raisonnement est celui qui donne lieu à l'action ou l'intention, alors que le second type est celui qui donne lieu aux croyances :

"We can distinguish theoretical reasoning, which most directly affects beliefs, from practical reasoning, which most directly affects plans and intentions. The upshot of theoretical reasoning is either a change in beliefs or no change, whereas the upshot of practical reasoning is either a change in plans and intentions or no change."⁹⁰

Aussi, dans la mesure où les attitudes réactives ne sont pas des croyances, elles appartiennent à la catégorie pratique, plutôt que théorique. Nous ne pouvons donc avoir, sur la question de l'abandon des attitudes réactives, que des raisonnements pratiques. Typiquement, ces raisonnements portent sur les avantages et inconvénients relatifs à nos choix et projets. Or, un tel raisonnement mènerait nécessairement à la conséquence que nous *devons* raisonnablement maintenir nos attitudes réactives.

⁸⁹ Kolodny, N. et Brunero, J. (2013). *Instrumental Rationality*. The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.) Cette définition de la raison pratique est appelée « rationalité instrumentale ».

⁹⁰ Harman, G. (1999). *Rationality', Reasoning, Meaning, and Mind*. (Oxford, 1999; online edn, Oxford Academic, 1 Nov. 2003), <https://doi.org/10.1093/0198238029.003.0002>

Cependant, je ne suis pas certain que cette réponse soit suffisante et je propose deux objections :

1. Il est faux que le déterminisme n'est pas le bon type de raison pour abandonner les attitudes réactives.
2. Le raisonnement pratique nécessaire à la thèse normative constitue un cas de *wishful thinking*.

Tout d'abord, nous avons vu, dans la section (1.3.), qu'Aristote identifiait deux conditions de la responsabilité morale qu'étaient l'absence de contrainte et l'absence d'ignorance. J'avais alors montré que ces deux conditions faisaient consensus, dans la mesure où il est évident qu'aucune théorie ne peut soutenir que l'on peut être moralement responsable, tout en étant contraint ou ignorant. Bien que la réflexion de Strawson ne concerne pas de ces questions, je pense néanmoins qu'il traiterait la contrainte et l'ignorance comme des motifs d'excuses. En effet, un agent ne peut pas faire preuve de mauvaise volonté s'il est ignorant, ou s'il est contraint par une force extérieure. Strawson en déduirait alors, que la colère suscitée par un acte contraint ou ignorant disparaît naturellement, à partir du moment où l'excuse est formulée. Cependant, il est important de se demander, plus précisément, ce qui empêche les agents de manifester de la mauvaise volonté, dans les cas d'ignorance et de contrainte. Quelle est la propriété inhérente et commune qui excuse la responsabilité morale ?

Je crois que la réponse à cette interrogation est la suivante : si un agent *ne peut pas s'empêcher d'agir tel qu'il le fait* (sous l'effet de la contrainte ou l'ignorance), alors il n'exprime pas de mauvaise volonté. Et s'il n'exprime pas de mauvaise volonté, alors il n'est pas responsable.

Cela dit, le déterminisme peut, au même titre que l'ignorance et la contrainte, être compris comme une situation dans laquelle *personne ne peut s'empêcher de faire ce qu'il fait*. À ce stade, on saisit mieux dans quelle mesure le déterminisme n'est pas seulement une doctrine générale, mais peut directement entraîner des conséquences pratiques. De ce point de vue, rien n'interdit même le déterminisme de constituer un motif d'excuse universel : si le déterminisme est vrai, personne ne peut exprimer de mauvaise volonté et personne ne peut être moralement responsable.

Dès lors, si ce raisonnement est correct, la prémissse selon laquelle le déterminisme n'est pas pertinent est certainement fausse. Or, si cette prémissse est fausse, la tension entre raisonnements pratiques et théoriques apparaît à nouveau.

Par ailleurs, une autre stratégie intéressante pourrait être de retourner la thèse normative contre elle-même. Si les raisonnements théoriques donnent essentiellement lieu à des implications théoriques, et les raisonnements pratiques essentiellement à des conséquences pratiques, Alors il est vrai qu'aucun raisonnement théorique ne doit donner lieu à des conséquences pratiques⁹¹. Mais il est tout aussi vrai qu'aucun raisonnement pratique ne doit donner lieu à des implications théoriques.

Or, les défenseurs de la thèse normative semblent oublier cette deuxième conséquence. Selon eux, par raisonnement pratique, nous devrions choisir de sauver nos attitudes réactives.

Néanmoins, les attitudes réactives ne sont pas de simples pratiques ou de simples réactions corporelles. Au contraire, comme nous l'avons vu, les attitudes réactives sont cognitivement complexes et nécessitent toujours une croyance en la responsabilité morale.

Dès lors, sur la base d'un raisonnement pratique, nous sauvegardons, non seulement nos attitudes réactives, mais encore nous croyons en la responsabilité morale de façon illégitime. On peut, ici, formuler la critique inversée, à savoir qu'un raisonnement pratique ne doit pas donner lieu à une croyance en la responsabilité morale. En effet, à supposer que le raisonnement pratique ait ce pouvoir-là, ce seraient nos intentions qui fixeraient nos croyances. Or, l'ajustement des croyances sur les intentions occasionne un type d'irrationnalité que l'on appelle *Wishful thinking* :

Imaginons, un cas : *un riche militant platiste, lors d'un débat mal engagé avec des scientifiques, décide d'acheter ses opposants en proposant une grosse somme d'argent pour tous ceux qui acceptent de croire véritablement en la platitude de la terre.*

Dans ce cas, il y a fort à parier que certains scientifiques ont, en considérant toutes les possibilités représentées par une telle somme, l'intention claire de croire au platonisme. Cette intention est le résultat d'un raisonnement pratique qui voit la croyance comme un moyen de parvenir à une fin souhaitable. Cependant, si l'un des scientifiques parvenait à croire ce qu'il a l'intention de croire, alors il serait irrationnel. En effet, son intention ne fournit pas une justification suffisante pour croire en la platitude de la terre. Malgré les avantages pratiques à tirer, seul un raisonnement théorique est en mesure de fournir une telle justification, car nos croyances répondent essentiellement à des raisons épistémiques. Les scientifiques sont donc psychologiquement incapables de croire en la platitude de la terre, même en échange de plusieurs millions.

Par analogie, la thèse naturaliste constitue un cas de *Wishful thinking*. En effet, la thèse nous demande de croire en la responsabilité morale, parce que nous avons l'intention de croire en la responsabilité morale. Or, cela est impossible. Si l'intention de croire en la responsabilité morale se justifie par nos raisons pratiques, la croyance en la responsabilité morale, quant à elle, peut uniquement se justifier

⁹¹ C'est à dire que le déterminisme ne doit pas intervenir dans les raisonnements autour des attitudes réactives.

par des raisons épistémiques. Nous sommes donc incapables de croire en la responsabilité morale sur la base des avantages que cette croyance représente, car ce sont nos croyances qui doivent déterminer nos intentions et non l'inverse.

Ainsi, pour ces deux raisons, la thèse normative n'est finalement pas suffisante pour sauver nos attitudes réactives du scepticisme.

3.5 Vers L’Illusionnisme

Je viens de défendre que les deux thèses de Strawson ne parviennent pas à résoudre le problème des attitudes réactives. Ainsi, la question demeure : comment expliquer notre incapacité à abandonner nos attitudes réactives et comment résoudre la tension entre raisonnement pratique et théorique ? La solution que j’aimerais suggérer, en dernier lieu, est celle de l’illusionnisme. Selon cette théorie, nous devons reconnaître le rôle important que jouent les illusions au niveau de nos pratiques de la responsabilité morale. Mais qu’est-ce que l’illusion et en quoi permet-elle de sauver les attitudes réactives ?

Selon Smilansky, l’illusion est une fausse croyance que nous tenons pour vraie. Elle constitue donc une réponse à la première question : nous sommes incapables d’abandonner nos attitudes réactives parce que nous vivons dans l’illusion de la responsabilité morale, c'est-à-dire que nous la croyons vraie, même si elle n’existe pas.

Par ailleurs, l’illusionnisme est non seulement une thèse descriptive (c'est un fait que les gens vivent dans l’illusion de la responsabilité morale), mais elle est aussi une thèse normative (l’illusion est souhaitable et doit être maintenue). Dès lors, la tension entre raisonnements pratiques et théoriques est résolue. En effet, même si le libre arbitre n’existe pas, son illusion suffit à éviter les conséquences pratiques de son inexistence. L’illusion permet la pratique des attitudes réactives et des attributions de responsabilité et résout ainsi le deuxième problème.

Néanmoins, cette thèse doit nécessairement faire face à une critique flagrante. Est-il rationnel d’avoir une croyance que l’on sait fausse ?⁹² Et si non, dans quelle mesure l’illusion est-elle plus qu’une simple erreur ?

Bien évidemment, l’illusion est irrationnelle, mais on pourrait néanmoins répondre que le propos n'est pas là. En effet, l'objection présuppose qu'il est possible de corriger ce qui est irrationnel. Il est, donc possible de corriger l’illusion. Or, je crois que la thèse de Smilansky est plus subtile et ne conçoit pas l’illusion comme un phénomène sur lequel nous avons un contrôle, ni un devoir de justification. L’illusion, contrairement à l’erreur, ne cède pas devant les preuves contraires et se maintient au sein de nos pratiques. Elle est une erreur qui ne se corrige pas. Selon mon interprétation, le philosophe veut effectivement montrer que cette irrationalité est fondamentalement enchevêtrée dans nos pratiques liées à la responsabilité morale :

⁹² Voir la notion de Wishful thinking développée dans la section précédente.

"It is not that we find out the truth and then say "Let's keep quiet about this", but that illusion is intimately entangled with our free-will-related beliefs, reactions and practices."⁹³

De ce point de vue, il existe une union forte entre l'agent et la responsabilité morale qui maintient l'illusion ; de la même manière qu'un père pourrait toujours croire en l'innocence de son fils, malgré les preuves évidentes de son crime. Ce que je cherche à montrer par cette analogie, est que l'illusion n'est pas seulement l'erreur du père qui refuse les preuves, mais elle aussi ce qui constitue le lien avec son fils. En effet, l'irrationalité du père est quasiment nécessitée par le lien qu'il entretient avec son fils. Sur cette base, on peut concevoir le lien entre l'illusion et la responsabilité morale de façon similaire. Si mon interprétation n'est pas erronée, il est même finalement possible de concevoir l'irrationalité de l'illusion comme incontournable dans le sens strawsonnien du terme.

Par ailleurs, on pourrait aussi répondre que cette thèse n'est pas irrationnelle, dans la mesure où l'illusion ne consiste pas nécessairement en une auto-persuasion vicieuse qui nous forcerait à croire ce que nous tenons pour faux. Au contraire, l'illusion est déjà présente et elle représente un avantage certain pour nos vies. L'illusionnisme considère donc simplement qu'il ne faut pas chercher activement à briser notre attachement aux attitudes réactives. Ainsi, Selon Smilansky, l'illusionnisme n'est pas non plus la thèse selon laquelle il faut diffuser une fausse croyance au libre arbitre, tel un noble mensonge. En effet, une telle manœuvre n'aurait pas conscience du fait que l'illusion est déjà largement présente en nous tous. Il s'agit donc simplement d'en accepter les avantages :

"It is not claimed that we need to induce illusory beliefs concerning free will, or can live with beliefs we fully realize are illusory. Rather, my claim is that illusory beliefs are in place, and that the role they play is largely positive. Humanity is fortunately deceived on the free will issue, and this seems to be a condition of civilized morality and personal value."⁹⁴

En ce sens, il ne faut pas se représenter l'illusion comme la contradiction constante de la réalité des choses. Par l'illusion, nous ne commençons pas à croire à ce que nous savons faux, mais nous continuons à croire à ce qui est faux. L'illusion n'est pas un mensonge. Nous abritons en permanence des croyances compatibilistes et incompatibilistes que l'illusion parvient à faire cohabiter, en ne confrontant pas leurs incohérences. Dès lors, l'illusionnisme ne prétend pas que l'illusion rende vraies nos croyances compatibilistes, mais simplement que, dans la mesure où nous ne sommes pas toujours conscients de nos engagements, nous aurions tort de vouloir toujours les clarifier.

⁹³ Smilansky, S. (2001). *Free Will: From Nature to Illusion*. Proceedings of the Aristotelian Society, 101, 71–95.

⁹⁴ Smilansky, S. (2001). *Free Will: From Nature to Illusion*. Proceedings of the Aristotelian Society, 101, 71–95.

Par l'illusion, nous sommes ainsi capables de sauvegarder les avantages de ces deux types de croyances : les croyances compatibilistes illusoires nous permettent d'être les participants de nos relations interpersonnelles, tandis que les croyances incompatibilistes tempèrent nos réactions trop véhémentes :

"However, some awareness of deterministic elements can be useful, mitigating resentment of others or self-recriminations. Illusion allows us the advantages of the libertarian picture together with the mitigating element, without full awareness either of the incoherence of the libertarian picture or of the contrariness of the compatibilist and ultimate perspectives."⁹⁵

En somme, la thèse illusionniste apporte une réponse précieuse à notre deuxième problème : bien que nos attitudes réactives soient rationnellement incompatibles avec le scepticisme, l'illusion nous permet, en pratique, de sauvegarder convenablement nos attitudes réactives.

Finalement, il est intéressant de noter que les avantages de cette position reposent, en un sens, sur sa faiblesse philosophique. C'est, en effet, l'irrationalité de nos croyances qui en sauve les pratiques. Cela paraît étrange. Néanmoins, je doute fortement que nous ayons des ressources supplémentaires pour concilier scepticisme et attitudes réactives :

"The result is not philosophically neat, but that, after all, is its merit: the original reality was that we face practical dangers if we try to make our (incoherent or contradictory) conceptions too clear, but that we ought not to give any of them up entirely. Illusion, in short, allows us to have "workable beliefs"."⁹⁶

Comme Smilansky, je suis conscient que l'illusionnisme n'est pas une thèse élégante, mais son aspect anti-théorique en fait peut-être sa force.

⁹⁵ Ibid.

⁹⁶ Ibid.

Conclusion

En conclusion, je soutiens que la thèse de Smilansky réussit à résoudre le problème des attitudes réactives. Plus encore, l'illusionnisme résout aussi les autres problèmes que j'ai évoqués dans ce travail, à savoir ceux du compatibilisme, du libertarianisme et du déterminisme dur. Tout d'abord, le type de libre arbitre qui est illusoirement tenu pour vrai est libertarien, ce qui évite les problèmes du compatibilisme. Ensuite, l'illusionnisme évite les problèmes métaphysiques du libertarianisme en soutenant que la responsabilité morale n'existe pas. Finalement, l'illusionnisme contourne, comme nous venons de le voir, les problèmes pratiques du déterminisme dur en établissant premièrement le caractère omniprésent des illusions dans nos vies morales et, deuxièmement, l'importance de la croyance en un libre arbitre et des attributions de responsabilité morale.

Néanmoins, l'illusionnisme a ceci de particulier qu'elle est une thèse philosophiquement repoussante. Dans mes développements, je suis d'abord parvenu à la conclusion que la colère était impossible et irrationnelle. Or, il a semblé ensuite que cette irrationalité était elle-même irrationnelle, dans la mesure où le problème se résout tous les jours dans nos pratiques. Finalement l'illusionnisme semble proposer de résoudre l'irrationalité de nos pratiques, en théorisant la pratique de l'irrationalité. Dès lors, il ne fait plus aucun doute que nous sommes arrivés à ce stade étrange où la théorie pense contre elle-même. J'avoue ne pas être convaincu pleinement, mais je doute qu'il soit possible de faire mieux. Peut-être le problème des attitudes réactives appartient-il à ce genre particulier de problèmes auxquels il vaut mieux ne pas répondre : il se montre quand on tente de le résoudre, et se résout quand on ne le montre pas.

Bibliographie

- Alvarez, M. (2021). *P. F. Strawson, Moral Theories and 'The Problem of Blame': 'Freedom and Resentment' Revisited*. Aristotelian Society Supplementary, Volume 95 (1):183-203.
- Aristote. *Ethique à Nicomaque*. 1109b30 – 1111b5.
- Ayer, A. J. (1980). *Free Will and Rationality*. In van Straaten, R. (Ed.), *Philosophical Subjects* (pp. 211-224). University of California Press
- Chevarie-Cossette, S.-P. (2017), *Responsabilité morale (A)*, dans Maxime Kristanek (dir.). L'Encyclopédie philosophique.
- Chisholm, R. (1964). *Human Freedom and the Self*. In Robert Kane (ed.), *Free Will*. Blackwell.
- Clarke, R. (2003). *Libertarian Accounts of Free Will*. New York: Oxford University Press.
- Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. MIT Press.
- Deonna, J., Rodogno, R. et Teroni, F. *In Defense of Shame: The Faces of an Emotion* (2011; online edn, Oxford Academic, 19 Jan. 2012),
<https://doi.org/10.1093/acprof:oso/9780199793532.001.0001>
- Fischer, J. M., & Ravizza, M. (1993). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press
- Fischer, J.M and Ravizza M. (eds.) 1993: *Perspectives on Moral Responsibility*. Ithaca, NY: Cornell University Press.
- Guillon, J.-B. (2016). *Libre Arbitre (GP)*. Dans Maxime Kristanek (dir.), l'Encyclopédie philosophique.
- Harman, G. (1999). *Rationality', Reasoning, Meaning, and Mind*. (Oxford, 1999; online edn, Oxford Academic, 1 Nov. 2003), <https://doi.org/10.1093/0198238029.003.0002>
- Hoefer, C. *Causal Determinism*. The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), Zalta, E. et Nodelman, U. (eds)
- Honderich, T. *Compatibilism, Incompatibilism, and the Smart Aleck*. Philosophy and Phenomenological Research 56 (1996), pp. 855–62, at pp. 856–9; cf. A Theory of Determinism.
- Kane, R. (1996), *The Significance of Free Will*. New York: Oxford University Press.
- Kane, R. (1996). *Libertarianism, Causal Determinism, and Decision Theory*. Philosophical Topics, 24(1), 117-146.
- Kolodny, N. et Brunero, J. (2013). *Instrumental Rationality*. The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), Zalta, E. et Nodelman, U. (eds.)
- Levy, Neil (2011). *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford, GB: Oxford University Press UK.

- McCormick, K. (2022). *The Problem of Blame: Making Sense of Moral Anger*. Cambridge: Cambridge University Press. doi:10.1017/9781108907071
- McKenna, M. (2012). *Conversation and Responsibility*. New York: Oxford University Press.
- McKenna, M. S. (1998). *The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism*. *The Journal of Ethics*, 2(2), 123–142.
- McKenna, M.S et Coates, D. J. *Compatibilism*. The Stanford Encyclopedia of Philosophy (Fall 2021 Edition), Edward N. Zalta (ed.)
- McKenna, M.S. *A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument*. *Philosophy and Phenomenological Research*, vol. 77, no. 1, 2008, pp. 142–59. JSTOR, <http://www.jstor.org/stable/40041222>.
- Mele, A. R. (2006). *Free Will and Luck*. New York, US: Oxford University Press.
- Nagel, T. (1976). *Moral luck*. *Proceedings of the Aristotelian Society*, 50, 137-151.
- Nahmias, E., Morris, S., Nadelhoffer, T., et Turner, J. (2004). *The Phenomenology of Free Will*. *Journal of Consciousness Studies*, 11(7-8), 162–179.)
- O'Connor T. et Franklin C. *Free Will*. The Stanford Encyclopedia of Philosophy (Winter 2022 Edition), Zalta, E. & Nodelman, U. (eds.).
- Pereboom, D. (2001). *Living Without Free Will*. Cambridge University Press.
- Roberts, R. C. (2003). *Emotions: An essay in aid of moral psychology*. Cambridge University Press.
- Russell, P., & McKenna, M. (Eds.). (2008). *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's "Freedom and Resentment"*(1sted.).Routledge. <https://doi.org/10.4324/9781315583112>
- Shoemaker D. (2017). *Response-Dependent Responsibility; or, a Funny Thing Happened on the Way to Blame*. *Philosophical Review*, 126(4), pp. 481–527.
- Smilansky, S. (1962). *Free Will: From Nature to Illusion*. *Proceedings of the Aristotelian Society*, 101 (2001), 71–95.
- Spinoza, B. *Éthique*. (1677). Édition établie, présentée et annotée par Charles Ramond. Collection Folio Essais, Gallimard, 2010.
- Strawson, G. (1993). *Freedom and Resentment*. In John Martin Fischer & Mark Ravizza (eds.), *Perspectives on Moral Responsibility*. Cornell University Press. pp. 67-100.
- Strawson, G. (1994). *The Impossibility of Moral Responsibility*. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 75(1/2), 5–24. <http://www.jstor.org/stable/4320507>.
- Strawson, P. F. (1962). *Freedom and resentment*. *Proceedings of the British Academy*, 48, 1-25.
- Van Inwagen, P. (1983). *Essai sur le libre arbitre*. Vrin. (Traduction française de l'ouvrage "An Essay on Free Will").

- Vihvelin, K. (2003). *Arguments for Incompatibilism*. The Stanford Encyclopedia of Philosophy (Fall 2022 Edition), Zalta, E. et Nodelman U. (eds.)
- Watson, G. (1987). *Responsibility and the Limits of Evil: Variations on a Strawsonian Theme*. Schoeman (Ed.), Responsibility, Character, and the Emotions. Cambridge University Press
- Watson, G. (2014). *Peter Strawson on Responsibility and Sociality*. Oxford Studies in Agency and Responsibility, 2, pp. 15–32.
- Williams, B. (1981). *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press.
- Wolf, S. (1981). *The Importance of Free Will*. Mind, 90(358), 386-405.