

0.0.1. Résumé

La responsabilité morale existe-t-elle ? Si oui, quelle est sa nature ? Voilà des questions particulièrement complexes puisqu'elles se retrouvent au confluent de l'éthique et de la politique, mais aussi de la philosophie du droit et de la métaphysique (en particulier du problème du libre arbitre). Du fait que d'aucuns jugent que la responsabilité morale est affectée par l'ignorance, elle touche aussi l'épistémologie. Cette entrée ne couvre donc pas toute l'étude, même récente, de la responsabilité morale. D'une part, elle fait plutôt l'économie de l'histoire de la responsabilité morale sur la longue durée, où Kant, Hume et Aristote sont les figures les plus influentes. D'autre part, alors que la responsabilité morale est parfois une notion qu'on applique à la collectivité plutôt qu'à l'individu, seule la responsabilité individuelle sera abordée ici.

Cette entrée se déroule en trois temps. Elle propose premièrement de distinguer plusieurs *sens* ou *concepts* de la responsabilité morale. Même si cette taxonomie est controversée, elle est essentielle à la compréhension des débats que nous explorerons par la suite. La deuxième partie abordera la nature fondamentale et les fonctions principales de la responsabilité morale. En particulier, il sera question d'évaluer le lien entre la responsabilité morale et l'attribution légitime du blâme ou de la louange. La troisième partie sera intimement liée à la deuxième puisqu'elle se penchera sur les différentes conditions nécessaires à la responsabilité morale. La part du lion sera réservée à la notion de contrôle. Les conditions historiques de la responsabilité morale seront ensuite examinées. Nous laisserons de côté les conditions épistémiques de la responsabilité morale ; néanmoins, la littérature contemporaine sur l'ignorance est foisonnante.

Il faut garder en tête que les théories positives de la responsabilité morale ont été conçues en constant dialogue avec un certain danger sceptique. Il n'est pas rare qu'une condition jugée nécessaire à la responsabilité soit, en fin de compte, impossible à remplir. En ce sens, le travail des théoriciens de la responsabilité morale ressemble à celui des épistémologues : il faudrait (peut-être !) trouver une façon de satisfaire des exigences normatives complexes sans tomber dans le scepticisme. Les arguments sceptiques (détaillés dans l'entrée sur le libre arbitre) – qu'ils reposent sur l'idée que nous ne pouvons pas être *la source ultime de nos actions*, mais que nous serions tels des agents manipulés ; ou qu'ils dérivent plutôt de la thèse selon laquelle *nous ne pouvons pas faire autrement* – forcent parfois les différents théoriciens à mettre à mal certaines intuitions bien ancrées.

0.0.1. Table des matières

1. Une taxonomie de la « responsabilité morale »
2. Différentes conceptions de la responsabilité morale

a. Rétributivisme, scepticisme et conséquentialisme

i. Le rétributivisme

ii. Le scepticisme

iii. Le conséquentialisme

b. Le naturalisme de Strawson et les attitudes réactives

c. Volitionnisme et attributionnisme : à la frontière entre métaphysique et pratique

3. Les conditions de la responsabilité morale

a. Le contrôle

i. La capacité de faire autrement et les possibilités alternatives

ii. Interpréter les possibilités alternatives

iii. Les contre-exemples à la Frankfurt

iv. Trois réponses aux contre-exemples à la Frankfurt

b. Manipulation et conditions historiques (et externalistes)

i. Les arguments de manipulation

ii. Trois réponses aux arguments de manipulation

4. Conclusion

5. Bibliographie

1. Une taxonomie de la « responsabilité morale »

Comme l'a montré H.L.A. Hart, le terme de « responsabilité » est hautement polysémique :

Comme capitaine de navire, X était [1] *responsable* de la sécurité des passagers et de l'équipage. Mais lors de son dernier voyage, il se saoula tous les soirs et fut [2] *responsable* de la perte du bateau et de son équipage. La rumeur courut qu'il était fou, mais les médecins considérèrent qu'il était [3] *responsable* de ses actions. Tout au long du voyage, il se comportait de façon [4] *irresponsable*, et des accidents variés dans sa carrière confirmèrent qu'il n'était pas une personne [5] *responsable*. Il soutenait à tout vent que des tempêtes hivernales exceptionnelles étaient [6] *responsables* de la perte de son bateau, mais au cours des procédures judiciaires intentées contre lui, il fut trouvé criminellement [7] *responsable* de négligence et, séparément, dans les procédures civiles, il fut tenu [8] *responsable* légalement de la perte de vies et de propriétés. Il est toujours en vie et il est moralement [9] *responsable* de la mort de plusieurs femmes et enfants. (Hart 2008, 211, traduction libre, l'italique est de l'auteur)

Hart distingue lui-même quatre sens majeurs de « responsabilité » dans ce passage : la responsabilité causale (2, 6, 9), la responsabilité comme capacité (3 ; 4 et 5 sont plutôt

des exemples de *vertu*), la responsabilité comme lien légal (« *liability* ») (7, 8) et la responsabilité comme rôle (1). Parmi ces sens de « responsabilité », certains sont proprement moraux, d'autres non. Par exemple, la responsabilité causale n'est clairement pas morale dans le cas du vent, mais elle l'est dans le cas du capitaine (9). La responsabilité comme capacité est aussi morale dans le sens où elle est une condition de possibilité de la responsabilité comme lien. Quant à la responsabilité comme rôle, il s'agit de la responsabilité prospective : le responsable est celui qui devra réparer les dégâts. Cette exigence n'est pas forcément morale : l'attribution de la responsabilité peut simplement être un mécanisme aveugle de répartition des fardeaux économiques qui répond à des impératifs pratiques (comme lorsque quelqu'un doit rembourser à son voisin les dommages que son chien a causés sur son terrain).

En tenant compte des distinctions de Hart et en nous focalisant sur la question proprement morale de la responsabilité, nous pouvons distinguer les sens suivants :

1. **Une relation entre un sujet et un événement (ou un fait) qui justifie le blâme ou la louange.** Ce premier sens est celui qui a le plus attiré l'attention des métaphysiciens intéressés à la question du libre arbitre. Il rappelle la responsabilité causale de Hart, mais se limite aux cas moraux. C'est un sens *rétrospectif* de la responsabilité puisqu'il s'agit de la responsabilité pour des actions, omissions ou attitudes passées ou pour les conséquences de ces actions, omissions ou attitudes. C'est la responsabilité morale prise en ce sens qui a fait l'objet d'un scepticisme persistant depuis au moins Spinoza (dans *L'Éthique*, 1677). Même si la plupart de philosophes s'intéressent d'abord et avant tout à la responsabilité pour les *actions*, de nombreux travaux ont été exécutés sur la notion d'*omission* (Haji 1992; Fischer 1997; A. M. Smith 2005; Clarke 2014) et de *conséquences* (Fischer et Ravizza 1993; Braham et van Hees 2012; Van Inwagen 2017).
2. **Un fait modal, celui de l'obligation.** Lorsqu'une personne est responsable en ce sens, c'est qu'elle a des obligations (des « responsabilités »). C'est donc un sens essentiellement prospectif de la responsabilité. Lorsqu'une personne *assume ses responsabilités*, elle remplit ses obligations. Lorsqu'on dit qu'un parent est responsable de ses enfants, on veut dire qu'il a des obligations à l'égard de ses enfants que les autres adultes n'ont pas. En ce sens, la responsabilité comme obligation se rapproche de la responsabilité comme rôle dont parle Hart.
3. **Une capacité ou un ensemble de capacités.** Ce sens vient tel quel de la classification de Hart. De quel(les) capacité(s) est-il question ? Il peut par exemple s'agir de la capacité à recevoir des obligations. Il est naturel de dire qu'une personne n'a la capacité d'avoir des responsabilités que si elle peut (en général) les satisfaire. Cette capacité se développe vraisemblablement par degrés ; elle est quasiment absente chez les jeunes enfants et est présente chez la plupart des adultes sains du point de vue cognitif et comportemental. La fiction juridique de la « personne majeure » ne correspond donc pas exactement à cette capacité, bien qu'elle en soit une approximation utile. Sur ce point, comme sur la première notion de responsabilité, les responsabilités légale et morale diffèrent (voir Hart 2008).
4. **Une vertu.** C'est le sens que l'on retrouve dans des phrases comme « Tu ne t'es pas comportée de façon responsable ». Ce quatrième sens est celui d'une *vertu* (comme la sagesse, la prudence ou le courage) : être responsable

signifie de se comporter de la bonne manière, entre autres en assumant ses responsabilités (dans le deuxième sens). Généralement, on pense qu'une personne ne peut se comporter de façon responsable que si elle a la capacité d'avoir des responsabilités ou si, comme l'enfant ordinaire, elle a une capacité partielle ou en développement d'avoir des responsabilités.

5. **Une composante d'un rapport à autrui.** Il s'agit du sens de « responsable » dans « tenir responsable ». Celle-ci ne semble en effet pas correspondre exactement aux sens précédents, même s'il paraît assez naturel de tenir responsables ceux qui *sont* responsables (dans le premier sens). Tenir responsable contient peut-être aussi quelque chose de l'ordre de l'attribution d'obligations spéciales (dans le troisième sens).

Ces sens sont parfois difficiles à distinguer parce qu'ils sont connectés les uns aux autres. C'est peut-être la raison pour laquelle plusieurs philosophes opèrent moins de distinctions (p. ex. Korsgaard 1992, 326). Si une personne a la capacité d'avoir des responsabilités (3), elle a normalement des responsabilités (2). Si une personne satisfait ses responsabilités, elle se comporte de façon responsable (4), sinon on dira qu'elle est responsable (1) de ne pas avoir satisfait ses responsabilités (2) ; on pourra donc la tenir responsable en toute légitimité (5). Se comporter de façon responsable (4) suppose donc de remplir ses responsabilités (2). Habituellement, l'on tient une personne responsable (1) au regard de ses responsabilités (2), bien que l'on puisse être responsable d'actions morales surrogatoires. Une personne ne peut être responsable de ses actions (1) si elle n'a aucune capacité à assumer des responsabilités (3) (par exemple si cette personne est un enfant ou une personne avec un trouble de santé mentale sévère).

Même dans les contextes où il n'est pas clair que la notion de responsabilité soit proprement morale, ces différents concepts sont pertinents. Par exemple, la notion de *responsabilité ministérielle*, utilisée dans le parlementarisme britannique réfère d'abord au fait que le ministre est responsable de la gestion de son ministère au sens où il en a l'obligation (dans le deuxième sens) ; ensuite, au fait que le ministre doit répondre des actions de son ministère, y compris celles de ses fonctionnaires et donc qu'il soit approprié de leur demander de rendre des comptes ou de les tenir responsables (dans le cinquième sens) ; enfin, au fait que le conseil des ministres est responsable des actions des ministres particuliers (dans le premier sens) (voir p. ex. Assemblée nationale du Québec 2016).

Chacun de ces sens, mais aussi les liens entre ces sens, est à l'origine d'un lot de problèmes philosophiques. Les sens (1), (2), (3) et (5) seront d'ailleurs omniprésents dans cette entrée même si le premier en sera certainement la pièce de résistance.

2. Différentes conceptions de la responsabilité morale

a. Rétributivisme, scepticisme et conséquentialisme

La responsabilité morale comme relation entre un événement et un sujet n'est pas qu'une relation causale. Il y a une différence, du moins *prima facie*, entre le rapport qu'entretient l'ours avec la destruction d'une tente et le rapport qu'entretient l'humain avec la destruction de l'environnement de l'ours. La responsabilité est donc connectée à la morale, mais de quelle manière ?

Comme on l'a vu dans la section précédente, une connexion supplémentaire est habituellement tracée pour distinguer la simple responsabilité causale de la responsabilité morale dans le premier sens : une connexion entre la responsabilité morale et le blâme et la punition. Le blâme (et à plus forte raison la punition) en l'absence de responsabilité morale serait inapproprié ou injuste. Certains tracent la même connexion avec la louange (voir Caruso et Morris 2016) tandis que d'autres (p. ex. Smart 1961; Darwall 2006) s'y refusent.

Qu'est-ce que le blâme ? C'est une notion qui mériterait une entrée à elle seule ; malgré cela il vaut la peine de mentionner au passage trois théories. D'abord, le blâme peut être l'expression d'une simple évaluation négative (Hieronymi 2004; Arpaly 2006). Ensuite, le blâme peut être une évaluation négative qui s'accompagne de l'expression d'émotions comme la colère (Wallace 1994). Enfin, le blâme peut être un acte communicationnel à propos du fait qu'une relation a été endommagée (Scanlon 2008).

Dans la mesure où nous nous intéresserons surtout à la tradition philosophique selon laquelle le contrôle est nécessaire à la responsabilité morale, la deuxième définition du blâme semblera la plus naturelle à adopter dans le cadre de cette entrée. En effet, la punition comme la communication d'émotions comme la colère sont des fardeaux que nous faisons porter à ceux qui en sont l'objet. Or, il semble naturellement plus juste de la limiter aux actions et omissions que l'agent contrôle. (Pour simplifier, sans contrôle, pas de responsabilité ; sans responsabilité pas de blâme et sans blâme pas de fardeau parce que sans contrôle, pas de fardeau.)

Comment la responsabilité morale assure-t-elle ce lien entre la pratique concrète du blâme (et peut-être aussi de la punition, de la louange et de la récompense) ? Autrement dit, qu'est-ce qui justifie cette pratique à l'endroit des gens jugés responsables ? On distingue initialement au moins quatre réponses :

- **Le rétributivisme** soutient que la responsabilité morale est ce qui fait qu'un agent *mérite* le blâme. La pratique du blâme est donc justifiée par des considérations de justice.
- **Le scepticisme** affirme que la responsabilité morale n'existe pas et donc que la pratique du blâme n'est pas justifiée.
- **Le conséquentialisme** maintient que la pratique du blâme est justifiée par ses conséquences et donc que sont responsables ceux qu'il serait *utile* de blâmer.
- **Le naturalisme** (strawsonien) déclare que la pratique du blâme n'a pas à être justifiée en général (cela n'implique rien sur la justification du blâme dans des cas particuliers).

Nous exposerons d'abord sommairement les trois premières réponses. Elles ont une histoire riche qui remonte à Aristote (*ÉN*, 1109b30-1111b4). En effet, Aristote argumente contre la réponse sceptique en y opposant aussi bien des considérations rétributivistes que conséquentialistes (les commentateurs s'opposent sur ce point). Ensuite, nous explorerons la réponse naturaliste et le travail sur les attitudes réactives de Peter Frederick Strawson qui, s'ils sont plus récents, ont néanmoins eu un impact majeur sur les débats contemporains.

i. Le rétributivisme

Une première réponse à la question de la justification du blâme est *rétributiviste* et donc *rétrospective*. Lorsqu'un agent est moralement responsable d'une action, d'une omission ou de leurs conséquences, cela signifie qu'il *mérite pro tanto* d'être blâmé (par lui-même par le truchement du remord (Clarke 2016) ou par autrui) si l'action était mauvaise et louangé si l'action était bonne. Ce mérite serait « basique » (« *basic desert* » (Pereboom 2001, 2014; Haas 2013)) :

Le sens de la responsabilité morale, qui est central au débat sur le libre arbitre, est mis en relief par la notion de *mérite de base* (Feinberg 1970; Pereboom 2001, 2007; G. Strawson 1994; Fischer 2007, 82; Clarke 2005; Scanlon 2013). Dire qu'un agent est moralement responsable d'une action dans ce sens signifie que cette action est sienne de telle façon que la personne mériterait d'être blâmée si elle comprenait que ce qu'elle avait fait était moralement mal, et qu'elle mériterait d'être louangée si elle comprenait que cette action était moralement exemplaire. Le mérite dont il est question est basique dans le sens où l'agent mériterait d'être blâmé ou louangé pour la seule raison qu'il est l'auteur de l'action, considérant qu'il en comprend le statut moral, et non, par exemple, en vertu de considérations conséquentialistes ou contractualistes. [...] La croyance qu'un agent est moralement responsable dans ce sens accompagne typiquement l'expression d'attitudes réactives comme le ressentiment moral ou l'indignation ; elle est donc rapprochée de la notion de responsabilité morale que P. F. Strawson met de l'avant (1962). (Pereboom 2014, 2, traduction libre)

Il faut noter qu'il n'y a en principe rien de contradictoire à dire qu'une personne est moralement responsable d'un événement (une action, une omission ou leurs conséquences) même si cet événement n'est ni bon ni mauvais (Pereboom 2014, 2). Néanmoins, une question demeure : quel est l'intérêt de parler de responsabilité proprement *morale* pour ces actions, omissions ou conséquences, si ce n'est que parce qu'il est possible qu'elles s'avèrent être morales ou immorales ultérieurement ?

Il est difficile de définir cette notion de mérite de façon entièrement satisfaisante. Par exemple, Dana K. Nelkin (2016, 177-78) rejette différentes définitions du mérite, qu'elles soient axiologiques (par exemple, une personne mérite d'être punie lorsqu'il est *bon* de la punir ou la blâmer) ou déontiques (par exemple, une personne mérite d'être punie lorsqu'il faut la punir ou la blâmer). Elle retient que la responsabilité morale dans le sens du mérite est une « raison conditionnelle de sanctionner » (Nelkin 2016, 179), ce qui réduit le mérite à peau de chagrin.

En revanche, certaines images évocatrices peuvent faire surgir de façon plus convaincante la notion de mérite. Une première illustration de la notion de mérite est l'idée selon laquelle on pourrait en principe justifier la damnation éternelle et la récompense du paradis (G. Strawson 1994) ou, plus humblement, la simple punition ou récompense ponctuelle après la mort (Caruso et Morris 2016) (puisque le paradis ou l'enfer n'ont pas de justification prospective en général et conséquentialiste en particulier). Une deuxième façon d'illustrer la notion de mérite (« *desert* ») est de la distinguer du fait qu'il soit simplement approprié ou adéquat (la « *fittingness* » ou adéquation) de ressentir un remords. Un passage frappant de *Moral Luck* de Bernard Williams est utile :

Le conducteur de camion qui, sans que ce soit sa faute, frappe un enfant se sentira différemment de tout spectateur, même si ce spectateur est assis à ses côtés dans la

cabine, sauf peut-être dans la mesure où ce spectateur pensera avoir pu lui-même prévenir l'accident ; c'est la pensée d'un agent. Nul doute, et à raison, les gens essaieront, pour reconforter le conducteur, de le tirer de son état sentimental, de le tirer en effet de là où il est vers quelque chose comme la place du spectateur, mais il est important que cela soit vu comme quelque chose qui doit être accompli, et de fait certains doutes seraient émis à propos d'un conducteur qui serait amené à cette position de spectateur trop banalement ou avec trop peu d'hésitations. Nous sommes désolés pour le conducteur, mais ce sentiment coexiste avec, en fait présuppose, quelque chose de spécial à propos de la relation de cet événement, quelque chose qui ne peut pas être simplement éliminé par la considération que ce n'est pas sa faute. (Williams 1981, 28, traduction libre)

Bernard Williams nous donne l'exemple du camionneur dans le but d'illustrer que des aspects fondamentaux de nos vies morales sont contaminés par la chance. Néanmoins, ici nous pouvons utiliser cet exemple pour illustrer la notion de mérite de base. Si l'on compare le conducteur de camion avec un second conducteur qui aurait été imprudent ou mal intentionné, on comprend que, même s'il est *approprié* pour les deux conducteurs d'éprouver de la culpabilité, seul le second conducteur *mérite* (scepticisme mis à part) d'éprouver de la culpabilité (Clarke 2016). Cette réponse signifie que la responsabilité morale a à voir avec la *justice* (Feinberg 1970). Une personne est moralement responsable d'une action, d'une omission ou de leurs conséquences lorsqu'il est *juste* de louer ou blâmer la personne pour cela. Bien sûr, il est possible que le premier conducteur de camion doive exprimer ce sentiment de culpabilité devant les parents (Jacobson 2013). Il a cette *responsabilité* (où responsabilité est synonyme d'obligation). Mais cela ne le rend pas *responsable* dans le premier sens.

ii. Le scepticisme

Cette première réponse ouvre la porte à une seconde proposition quant au lien entre la responsabilité morale et le blâme, cette fois-ci *sceptique*. En fait, la réponse est presque identique à celle du rétributiviste au sens où l'on postule que la responsabilité morale est ce qui fait mériter le blâme ou la louange (Haas 2013; Caruso et Morris 2016). La différence fondamentale est que le sceptique croit qu'il n'existe pas une telle chose que le mérite dont il est question. Par conséquent, il soutient que la pratique de l'attribution de la responsabilité morale en général et celle du blâme en particulier (mais aussi de la louange, de la punition et de la récompense) soient abrogées.

Pourquoi ? Si la responsabilité morale est ce qui fonde le caractère *juste* du blâme (et peut-être de la punition, de la louange et de la récompense), il semble que certaines conditions doivent être satisfaites pour qu'un sujet soit responsable d'une action. Autrement dit, le mérite a ses conditions.

Le sceptique suggère donc d'admettre l'argument suivant :

Contrôle : Pour être responsable d'une action ou d'une omission, un sujet doit avoir exercé un contrôle suffisant sur cette action ou omission.

Scepticisme du contrôle : Nous n'exerçons jamais le contrôle nécessaire pour être responsables de nos actions ou omissions.

Scepticisme de la responsabilité morale : Nul n'est responsable de ses actions ou omissions.

Cela donne lieu à des doctrines nous intimant d'abandonner l'attribution de la responsabilité et ses pratiques connexes comme le blâme et la louange (Pereboom 2001, 2014, Waller 1990, 2014; Caruso forthcoming; G. Strawson 2010). Un argument similaire est parfois présenté pour montrer comment l'absence de contrôle est une menace pour la responsabilité dans le sens des obligations (Haji 2016). Cela suggère des problèmes de conciliation entre le scepticisme de la responsabilité morale et l'éthique des droits (Nelkin 2014).

Devant ce constat, les sceptiques proposent des solutions différentes. Certains suggèrent d'amender en profondeur nos institutions et nos pratiques de manière à remplacer la punition et le blâme par des solutions que nous utilisons déjà dans des cas où nous n'attribuons pas de responsabilité morale. Ainsi, Derk Pereboom et Gregg D. Caruso défendent un *modèle de la quarantaine* et une *révision de la pratique du blâme* (Pereboom 2001, 2008, 2013a, 2014; Caruso forthcoming). Le modèle de la quarantaine propose de remplacer les punitions institutionnelles par des solutions analogues à celles mises en place pour prévenir la contamination d'infections par le truchement de la quarantaine. Si les criminels ne méritent pas le blâme ou la punition, cela ne signifie pas qu'il serait adéquat de les laisser en liberté. Leur emprisonnement est justifié par le droit à l'autodéfense, et donc tant et aussi longtemps que les criminels constituent une menace.

Quant au remplacement du blâme, il ne s'agit pas non plus de fermer les yeux sur les torts d'autrui ou sur nos propres torts. Par exemple, Pereboom suggère de remplacer les réactions émotives qui constituent en partie le blâme par une forme de tristesse morale, c'est-à-dire la tristesse d'avoir mal agi ou d'être le genre de personne qui agit mal (dans le cas du remord) ou encore la tristesse que quelqu'un d'autre ait mal agi ou soit le genre de personne qui agit mal (dans le cas du blâme). Bien sûr, s'il n'y a pas de différence entre blâme et évaluation critique (Hieronymi 2004; Arpaly 2006), cette suggestion est minimale ou vide.

iii. Le conséquentialisme

Une troisième réponse à la question du lien entre la responsabilité morale et du blâme ou de la louange est *conséquentialiste* et donc *prospective* (Smart 1961, 1973, chap. 7; Dennett 1984; Vargas 2005, 2009, 2015). La question du mérite est complètement évacuée :

De qui était-ce *la* responsabilité ? L'utilitariste de l'acte [incluant Smart] répondra de façon cohérente que la notion de *la* responsabilité est un absurde morceau de métaphysique qui devrait être remplacée par « Qui serait-il utile de blâmer ? » (Smart 1973, 54, traduction libre)

En un sens, le conséquentialiste est d'accord avec le sceptique puisqu'il croit que la pratique du blâme n'est *pas justifiée à cause du mérite des agents*.

C'est l'*attribution* de la responsabilité morale plutôt que la responsabilité morale elle-même qui est importante pour le conséquentialiste. Il ne faut l'attribuer que lorsqu'il y a espoir de changer le comportement de la personne que l'on blâme (et donc, à qui l'on

attribue la responsabilité d'une action, d'une omission, d'un trait ou d'une attitude négative). C'est donc que, mis à part les cas où l'on blâme quelqu'un pour changer les comportements des *autres*, il ne faut blâmer que ceux qui ont les capacités de contrôle suffisantes pour modifier leur comportement.

Smart (et d'autres après lui, notamment Darwall 2006) notent une différence entre la louange et le blâme, quant à leur lien avec la responsabilité morale. La louange (« sincère », doit-on préciser) serait une simple forme de compliment qui viserait d'abord à décrire et ensuite à encourager (Smart 1961, 304). Nous pouvons louer un mathématicien pour son talent ou un footballeur pour sa course sans pour autant blâmer le mathématicien moins talentueux ou le footballeur qui est tombé un peu plus tôt (Smart 1961, 303). Fondamentalement, la louange n'implique pas toujours la responsabilité ; elle s'oppose donc plutôt à la critique (« *dispraisal* »). Le blâme, quant à lui, serait une forme de critique qui *s'accompagnerait* d'une attribution de responsabilité (Smart 1961, 305).

Si la responsabilité morale n'est qu'une chimère métaphysique, pourquoi parler de blâme et pas simplement de critique ? Smart le dit franchement : la notion populaire de responsabilité est fondée sur des incohérences métaphysiques implicites. Le blâme est donc en général irrationnel (Smart 1961, 306). Sans pour autant abandonner la notion de responsabilité morale et de blâme, Smart recommande de les délaissier, sauf lorsqu'ils sont utiles (par exemple, pour communiquer (voir Sie 2000, 2005)). C'est en ce sens que cette conception de la responsabilité morale est dite « conséquentialiste ». Il faut noter que les théoriciens conséquentialistes de la responsabilité morale sont conscients que leur conception de la responsabilité est différente de son acception courante ; c'est pourquoi certains d'entre eux se disent ouvertement « révisionnistes » (voir en particulier Vargas 2005, 2015).

b. Le naturalisme de Strawson et les attitudes réactives

Une dernière réponse à la question de ce qui justifie le blâme, la punition ou la louange pour les individus responsables est *naturaliste* (P. F. Strawson 1974, 1985, Russell 2008, 2013). Dans son influent *Freedom and Resentment* (Liberté et ressentiment), Peter Frederick Strawson soutient que nous *n'avons pas à justifier* la pratique générale du blâme et de la louange pour ceux que nous considérons comme responsables moralement (ce qui n'est peut-être pas le cas de la punition).

Pour comprendre cette réponse, il faut d'abord explorer rapidement la théorie *descriptive* que Strawson offre. Strawson rompt avec une tradition plus « théorique » pour s'intéresser à la *pratique* de l'attribution de la responsabilité morale (un peu comme Bernard Williams). Il note d'abord que nous avons, depuis toujours, adopté des attitudes différentes face aux individus responsables (dans le quatrième sens, c'est-à-dire comme personnes *majeures*) et non responsables. Ce sens de la responsabilité doit être le point de départ de notre réflexion sur la responsabilité en général pour Strawson. De fait, notre façon d'interagir avec les enfants, les animaux et les personnes atteintes de maladies mentales sévères d'une part et avec les adultes sains d'esprit ou dépourvus de déficiences majeures d'autre part est fondamentalement différente (P. F. Strawson 2008, 9). Nous avons, envers les premiers, une attitude *objective*, c'est-à-dire que ces individus sont l'objet de prévision et de contrôle. Nous les décrivons et les analysons d'une façon qui semble incompatible avec

l'évaluation morale (P. F. Strawson 1985, 40). Lorsque nous expliquons leurs actions, nous parlons en termes de causes biologiques ou psychologiques plutôt qu'en termes de raisons. (« Il a fait une crise parce que son taux de sucre était bas » ; « elle s'est précipitée sur le gâteau parce qu'elle avait trop faim ».) Ce que nous exigeons d'eux est moindre. Nous entretenons, avec les seconds (les adultes responsables), une attitude *subjective*, soit une attitude de sujet à sujet, teintée par des exigences interpersonnelles. Lorsque nous expliquons leurs actions, nous utilisons le vocabulaire des raisons. (« Il est allé à l'église pour faire plaisir à sa mère » ; « j'ai choisi un vélo en aluminium parce que je veux me rendre rapidement au travail », etc.)

La pratique sociale qui consiste à tenir des individus responsables de leurs actions est une pratique qui découle de cette distinction entre attitudes objectives et subjectives. Nous avons des *attitudes réactives* envers des individus responsables, alors que nous les suspendons (ou ne les avons tout simplement pas) lorsque ces individus ne sont pas responsables. Les attitudes réactives sont des réactions émotionnelles comme le blâme, la colère, le ressentiment, le pardon, l'amour, le regret, la fierté, le remords ou la honte (P. F. Strawson 2008, 5, 16) que nous adoptons à l'endroit d'individus qui ont manifesté de la bonne ou de la mauvaise volonté.

Strawson note qu'il nous arrive de remettre en cause ou de suspendre nos attitudes réactives, mais que cela s'effectue toujours de manière locale. Parfois nous suspendons nos attitudes réactives à cause d'une particularité du sujet (le sujet est un enfant) ; parfois nous les suspendons à cause d'une particularité de la situation (le sujet ignorait que ces paroles me feraient de la peine) ; parfois encore, nous suspendons nos attitudes réactives pour nous protéger, en adoptant une attitude objective (P. F. Strawson 1985, 34).

La distinction entre les deux premières catégories de suspension d'attitudes réactives est reprise par Gary Watson et est encore centrale aujourd'hui (voir par exemple Ebels-Duggan 2013, 153-54). Watson sépare d'une part l'*excuse*, par exemple lorsque je suspends mon ressentiment contre un ami qui m'a visiblement heurté parce qu'il ignorait que ce qu'il me disait était insultant ; et d'autre part l'*exemption*, comme lorsque je suspends mon ressentiment parce que je réalise que la personne qui m'a lancé une balle de neige est un enfant (Watson 2008, 2012). Il n'y a donc, dans la pratique, que des excuses ou des exemptions *locales* (P. F. Strawson 2008, 13-14).

Nous pouvons maintenant poser la question à nouveau : pourquoi ne doit-on pas justifier la pratique de la responsabilité morale ainsi que du blâme, de la louange et plus généralement de tout l'appareillage des attitudes réactives ? Strawson nous offre trois raisons :

- La première raison est typiquement humienne (P. F. Strawson 1974, 12-14, 1998, 242) : il serait 'pratiquement impensable', c'est-à-dire *impossible en pratique*, d'abandonner nos attitudes réactives (P. F. Strawson 2008, 12, 14, 1985, 33). Strawson traite alors l'idée sceptique de la même manière avec laquelle Hume analyse la pratique de l'induction : comme quelque chose qu'il est parfaitement inutile de questionner parce que trop ancré naturellement (Hume 1739, Traité de la nature humaine, Livre I, Partie IV, Section VII). Strawson admet tout de même que, prise seule, cette raison n'est pas déterminante (P. F. Strawson 1998, 242).

- La deuxième raison est typiquement wittgensteinienne (P. F. Strawson 1974, 12, 1998, 242) : la remise en question de la pratique de l'attribution de la responsabilité morale *n'a aucun sens* (P. F. Strawson, Sen, et Verma 1995, 274; Wittgenstein 1974, §35). Puisque le doute ne peut avoir un sens qu'à même un système (Wittgenstein 1974, §105), et que ce système est celui des attitudes réactives et de la responsabilité morale, tout doute qui se dirigerait contre la responsabilité morale elle-même (et donc contre tout ce système) se détruirait lui-même (voir Wittgenstein 1974, §63, §158, pour un argument similaire contre le scepticisme en épistémologie).
- La troisième raison rejoint le conséquentialisme : même si nous avons la possibilité de remettre en cause notre pratique, cela aurait des conséquences désastreuses sur la vie humaine, qui se retrouverait considérablement appauvrie. En effet, l'abandon des attitudes réactives et l'adoption permanente de l'attitude objective qui en découlerait nous condamnerait à l'isolement (P. F. Strawson 1974, 12). Il ne serait donc pas rationnel d'être sceptiques. On peut reformuler cette raison de façon kantienne : il y aurait deux perspectives, l'une théorique et l'autre pratique, et tout ce qui relève de la pratique devrait être justifié selon cette même perspective, qui serait en quelque sorte autonome ou « insularisée » (P. F. Strawson 1985, 37-40).

À ces critiques de la réponse sceptique, Strawson ajoute une critique de la réponse rétributiviste et de la réponse conséquentialiste. D'abord, la justification conséquentialiste suggèrerait que l'attitude objective est la seule valable, ce qui serait non seulement faux, mais « offensant pour les coupables (« *offensive for the offenders* ») (P. F. Strawson 2008, 4), en plus de « donner froid dans le dos » lorsque l'on considère la récupération autoritaire de cette idée (P. F. Strawson 2008, 23). Bien sûr, les conséquentialistes tentent de montrer qu'il est mal de punir les innocents puisque cela aurait à long terme une influence néfaste sur les gens. Mais cette justification semble être *du mauvais type* (P. F. Strawson 2008, 2, 22) : c'est directement *parce qu'ils sont innocents que nous ne punissons pas les innocents*. Pour sa part, la justification rétributiviste suppose une métaphysique « inintelligible » (P. F. Strawson 2008, 24) et « propre à engendrer la panique » (P. F. Strawson 2008, 27), c'est-à-dire l'idée selon laquelle nous serions la cause non causée de nos actions. Si certaines des critiques de Strawson peuvent aujourd'hui paraître datées ou caricaturales, c'est bien parce que ses contemporains ont pris acte de ses leçons (par exemple, des sceptiques comme Pereboom ont conçu des théories positives raffinées qui suggèrent de remplacer le blâme sans recommander l'adoption généralisée de l'attitude objective).

c. Volitionnisme et attributionnisme : à la frontière entre métaphysique et pratique

Strawson croit donc que la responsabilité morale est un élément nécessaire à la vie humaine, qui *ne peut être remis en cause* pour des raisons métaphysiques générales comme la vérité du déterminisme ; qu'il *n'y a pas de sens* à le remettre en cause pour ces raisons ; enfin, qu'il serait *dévastateur* de supprimer. Cela voudrait dire que l'existence de notre pratique d'attribution de la responsabilité morale et des attitudes réactives connexes serait immunisée contre des considérations métaphysiques d'ordre sceptique. Il n'en résulte pas pour autant qu'il n'y a pas de conditions métaphysiques nécessaires à la responsabilité morale. Simplement, si de telles conditions existent, elles doivent en principe être généralement satisfaites. Autrement dit, l'impossibilité du

scepticisme n'implique pas directement que la question de la responsabilité morale n'est pas une question métaphysique.

L'une des particularités principales du débat sur la responsabilité morale est précisément de se retrouver à la jonction entre la métaphysique et l'éthique et donc de dépendre des contributions des deux champs (Arpaly et Schroeder 1999, 161; D. Shoemaker et Tognazzini 2015, 1). Ainsi, la notion de responsabilité morale dans le premier sens abordé plus haut (celui d'une relation particulière entre un événement et un sujet) est fondamentalement (mais pas complètement) métaphysique. Ce n'est qu'une fois que la relation de responsabilité est établie que l'on se pose la question si l'action était bonne ou mauvaise, blâmable, digne de louanges, ou neutre ; ou s'il y avait des raisons externes de disculper le sujet (Fischer et Tognazzini 2011).

Au contraire, la notion de responsabilité dans le cinquième sens, celle qui est présente dans la locution « tenir responsable », comme celle de la qualité d'un rapport à autrui est essentiellement éthique. Si on entame notre enquête sur la responsabilité morale à partir de cette notion, alors la responsabilité dépend non pas fondamentalement de conditions métaphysiques de contrôle, mais des attitudes que nous avons l'obligation de manifester à l'égard des partenaires dans notre communauté morale (Scanlon 1998; A. M. Smith 2007).

Dès lors, il est possible de distinguer deux courants de pensée sur la responsabilité morale, même s'ils ne sont pas suffisamment unifiés pour qu'on leur donne le nom « d'école » (M. McKenna 2008, 30). La popularisation de cette taxonomie est due entre autres à Neil Levy (2005).

- **Le courant volitionniste** s'intéresse d'abord à la notion de choix et de contrôle individuel. Il postule que pour être responsable d'une action, il faut que cette action fasse partie de la sphère de notre contrôle. Une action ou une attitude ne peut donc être l'objet correct d'un blâme si elle est involontaire, à moins qu'elle soit le résultat prévisible d'une autre action volontaire. (Frankfurt 1971; Wolf 1993; 2002; Fischer 1994, 2006; Fischer et Ravizza 1998; Mele 1995; Ginet 1996; Kane 1996, 2016; Watson 1996; Pereboom 2001, 2014; Levy 2005; O. Brink et Nelkin 2013; Waller 2014; Sartorio 2016a; Clarke 2016; M. J. Zimmerman 2016)
- **Le courant attributionniste** s'intéresse d'abord à ce que nous nous devons les uns les autres (par exemple ce que deux amis se doivent) et voit dans la responsabilité morale une notion interpersonnelle entre deux agents qui font partie d'une même communauté morale. Les attributionnistes partent de la notion de *tenir responsable* pour expliquer celle d'*être responsable*. Le fait que nos attitudes ou nos jugements brisent les exigences légitimes de notre communauté ou de nos relations (p. ex. Ebels-Duggan 2013) ou le fait que certains nous tiennent responsables (p. ex. Korsgaard 1992) suffit à nous *rendre* responsables. *Prima facie*, rien n'indique que seuls les attitudes ou les jugements volontaires sont l'objet correct d'un blâme, puisque, parfois, nos attitudes ou jugements sont inappropriés (comme des oublis et fous rires ou comme certains jugements implicites racistes ou sexistes) même s'ils sont inconscients ou incontrôlables. (Smart 1973, 46; P. F. Strawson 1974; Korsgaard

1992; Williams 1993; Adams 1985; Scanlon 1998; A. M. Smith 2005, 2007, 2008; Kelly 2008, 274-75; Ebels-Duggan 2013; Cane 2016)

En général, les volitionnistes et les attributionnistes semblent postuler que la responsabilité morale a différents objets : les actions ou les omissions dans le premier cas et les attitudes dans le second.

L'opposition entre volitionnisme et attributionnisme semble souvent assez franche. Par exemple, d'un côté, un Derk Pereboom dit s'intéresser à la responsabilité morale, en tant qu'il est question de la capacité de *contrôle* nécessaire pour *mériter* dans le sens d'un *mérite de base* ('*basic desert*') et donc abstrait de considérations conséquentialistes et *contractualistes* (Pereboom 2014, 2, cité plus haut). D'un autre côté, la plupart des attributionnistes assoient leur théorie de la responsabilité morale sur une *forme de contractualisme* (Scanlon 1998; et Korsgaard 1992). L'attributionnisme est en effet assez courant chez les théoriciens du contrat social, car le contrat social suppose, pour réussir, des standards de coopération que chacun doit respecter. La notion première est donc celle de *devoirs*. Si nous ne remplissons pas nos devoirs, alors le contrat social justifie que nous distribuions le blâme et la responsabilité. Le blâme et la responsabilité sont distribués de façon semblable aux biens économiques, c'est-à-dire conformément à des principes établis dans un contrat social (Kelly 2008, 275). C'est une conception de la responsabilité qui est au service d'une moralité conçue comme outil pratique et donc qui tente de faire l'économie de la métaphysique (Rawls 1980; Kelly 2008, 275).

Le problème majeur de la conception attributionniste est que ce que nous nous devons les uns aux autres (ou, autrement dit, les principes que nous adoptons dans le contrat social) semble dépendre de ce qui est en notre contrôle. Des sceptiques établiraient un contrat social différent de celui que des rétributivistes voudraient instaurer. Si tel est le cas, alors il vaut peut-être mieux commencer l'analyse avec les volitionnistes et donc formuler une théorie du contrôle. Il faudrait alors d'abord établir une théorie de la responsabilité morale dans le premier sens (et seulement plus tard dans le cinquième) :

[...] Le fait qu'une personne *P* soit rétrospectivement moralement responsable pour une chose *x* dépend, non pas de s'il est approprié, en fait ou en principe, pour une personne actuelle d'adopter des attitudes réactives à l'égard de *P* par rapport à *x*, mais cela dépend plutôt de si *P* mérite ou est à la hauteur d'une telle réaction ; et cela en retour a uniquement à voir avec *P* et rien à voir avec qui que ce soit d'autre qui se trouve à exister. (D. Shoemaker et Tognazzini 2015, 24, qui résument la thèse de Zimmerman, traduction libre)

La question de s'il est approprié de tenir responsable serait une question se poserait *après* avoir déterminé si la personne est moralement responsable (voir p. ex. Todd forthcoming; Fischer et Tognazzini 2011). Ainsi, même si un bandit est moralement responsable d'un braquage de banque, son collègue qui a braqué la banque avec lui n'est vraisemblablement pas en position de le tenir responsable et de le blâmer.

Malgré l'opposition entre le volitionnisme et l'attributionnisme sur l'objet de la responsabilité morale et sur la condition de contrôle, certains théoriciens (p. ex. M. McKenna 2008; Sripada 2015; Nelkin 2016) ont tenté un rapprochement entre les deux

courants, par exemple en diminuant, mais sans toutefois éliminer, la condition de contrôle nécessaire à la responsabilité.

La complexité de la question de la responsabilité morale a donné naissance à d'autres taxonomies. Certains proposent, par exemple, de distinguer les théories de la responsabilité morale par la *fonction* qu'elles donnent à la notion de responsabilité morale. Est-il d'abord question de *sanction*, de *jugement évaluatif* ou de *communication* (Cogley 2013)? Si la littérature s'est concentrée sur les deux premières fonctions du blâme, certains ont exploré l'idée que lorsque nous blâmons, nous communiquons nos attentes (Sie 2000; Scanlon 2008). Nous exigeons également que les individus blâmés *rendent des comptes*, qu'ils *répondent de leurs gestes* (la notion difficilement traduisible d'« *answerability* » est importante (Watson 2004; D. W. Shoemaker 2011)).

Il est aussi courant de distinguer les théories *compatibilistes* et *incompatibilistes* de la responsabilité morale. Le compatibilisme est la thèse selon laquelle l'existence de la responsabilité morale est compatible avec le déterminisme. L'incompatibilisme est simplement la thèse selon laquelle la responsabilité morale est incompatible avec le déterminisme. Le déterminisme est défini de plusieurs manières : notamment comme la thèse selon laquelle tout événement, y compris nos décisions, a des causes suffisantes ; ou comme la thèse selon laquelle, la conjonction des lois de la nature et de la description complète du monde à un certain moment implique nécessairement la description complète du monde à tout autre moment. La distinction entre volitionnistes et attributionnistes place tous les incompatibilistes (mais pas seulement eux) dans le camp volitionniste puisque les arguments en faveur de l'incompatibilité entre responsabilité morale et déterminisme reposent sur l'idée que le déterminisme serait incompatible avec une certaine forme de contrôle.

3. Les conditions de la responsabilité morale

Rares sont les philosophes qui ont osé offrir des conditions nécessaires et suffisantes à la responsabilité morale. On n'a, pour le comprendre, qu'à penser à la diversité des phénomènes que sont la manipulation, la coercition, l'ignorance, la maladie, les phobies, les traumatismes et l'inexpérience pour saisir la difficulté de la tâche. À cela s'ajoute la possibilité d'exprimer les conditions de la responsabilité morale positivement (par exemple, l'action ou l'omission doit résulter d'une intention) et négativement (l'action ou l'omission ne doit pas résulter de la menace). Ce qui est généralement attendu d'une théorie de la responsabilité n'est donc pas tant une liste complète de conditions suffisantes et nécessaires que l'explicitation des conditions centrales et controversées nécessaires à la responsabilité morale (en général le contrôle), doublée de la présomption que les autres conditions ne causeront pas des difficultés et en particulier des défis sceptiques (telle l'exigence que le déterminisme soit faux) (Pereboom 2001, 100-101).

Autrement dit, une théorie de la responsabilité morale suppose une réponse à l'argument sceptique que nous avons précédemment examiné :

Contrôle : Pour être responsable d'une action ou d'une omission, un sujet doit avoir exercé un contrôle suffisant sur cette action ou omission.

Scepticisme du contrôle : Nous n'exerçons jamais le contrôle nécessaire pour être responsables de nos actions ou omissions.

Scepticisme de la responsabilité morale : Nul n'est responsable de ses actions.

Nous avons vu que la réponse des théories attributionnistes consistait à nier la première prémisse. Les théories volitionnistes, qui sont majoritaires et acceptent **Contrôle**, doivent donner une théorie du contrôle qui rend **Scepticisme du contrôle** fausse. Il leur faut donc élaborer une *théorie du contrôle*.

Nous examinerons à présent certaines de ces théories en travers l'exploration de scénarios difficiles et philosophiquement fructueux : les contre-exemples à la Frankfurt et les scénarios de manipulation. Ces scénarios permettent d'explorer plusieurs divisions dans les théories du contrôle, notamment concernant le caractère historique de la responsabilité et son lien avec la notion de capacité à faire autrement. Nous terminerons cette section en abordant la notion d'ignorance, laquelle pose un défi de taille dans l'élaboration des conditions nécessaires et suffisantes à la responsabilité morale. Il semble en effet que l'ignorance serve tantôt à disculper et tantôt à inculper (cf. Sartorio 2016b). Faute d'espace, il nous a fallu laisser en plan d'autres conditions discutées dans la littérature récente sur la question, telle que la conscience (p. ex. G. Strawson 1989; Arpaly 2003; Levy 2011, 2014; De Brigard 2015) ou l'absence de coercition (p. ex. Nozick 1997; Hyman 2015).

a. Le contrôle

La question du contrôle nécessaire à la responsabilité morale pour des actions, des omissions ou leurs conséquences (donc dans le premier sens distingué) est la « pièce de résistance métaphysique » des débats sur la responsabilité morale. Même si du côté des attributionnistes, où l'on admet volontiers que nous sommes responsables de certaines attitudes, jugements ou émotions (par exemple, l'oubli (A. M. Smith 2005), les préjugés sexistes (Ebels-Duggan 2013) et la colère (Adams 1985)), une forme générale de capacité à contrôler est nécessaire. C'est la raison principale pour laquelle les animaux, les jeunes enfants et les adultes souffrant des handicaps mentaux sévères sont d'ordinaire exclus de la sphère des êtres blâmables (cf. P. F. Strawson 1974; Fischer et Tognazzini 2011).

La capacité de contrôle nécessaire à la responsabilité morale reçoit parfois le nom de « libre arbitre » (p. ex. Pereboom 2001, 2, 2014; Mele 2006a; M. McKenna 2008; O'Connor 2016; Caruso et Morris 2016; van Inwagen 1983 suggère à ses lecteurs qui voudraient séparer libre arbitre et responsabilité morale de remplacer « libre arbitre » par « libre arbitre nécessaire pour la responsabilité morale »). Cependant, il faut faire attention puisque certains philosophes préfèrent admettre qu'une condition de contrôle est nécessaire à la responsabilité morale, mais que ce type de contrôle n'est pas aussi fort que le libre arbitre lui-même (Frankfurt 1971, 19; Fischer et Ravizza 1998; Fischer 2012). Ainsi Fischer se dit « semi-compatibiliste », c'est-à-dire qu'il est persuadé que le déterminisme est compatible avec la responsabilité morale, mais peut-être pas avec le libre arbitre. D'autres philosophes (p. ex. Waller 1990, 2014) tiennent aussi à démarquer la condition de contrôle nécessaire à la responsabilité morale et le libre arbitre, mais pour la raison inverse : il serait plus difficile de satisfaire les conditions nécessaires à la responsabilité morale que de satisfaire les conditions nécessaires au

libre arbitre. D'autres encore, tout en insistant sur l'importance du caractère volontaire des actions pour la responsabilité morale et donc d'une certaine forme de contrôle, croient que certaines notions derrière le libre arbitre (celle de volonté (« *will* »)) sont suspectes et donc que des concepts comme « volontaire » sont plus utiles que celui de libre arbitre (voir Hyman 2015). Nous renverrons donc les lecteurs intéressés par la question du libre arbitre à proprement parler à l'entrée qui porte ce nom dans la présente encyclopédie.

i. La capacité de faire autrement et les possibilités alternatives

La discussion contemporaine sur la responsabilité morale est initiée par la remise en question d'un principe que l'on attribue alors à la majorité du corps philosophique sous une forme ou une autre (p. ex. Aristote (ÉN), Hume 1748; Ayer 1954) :

Le principe des possibilités alternatives (PAP) : Un agent est responsable d'une action φ seulement si cet agent avait pu faire autrement que φ (y compris en omettant d'agir).

Trois remarques préliminaires s'imposent. Premièrement, tel que formulé, le PAP régle seulement les *actions* et non les omissions. Si certains auteurs en ont donné une analyse asymétrique (Fischer et Ravizza 1998, 125; Sartorio 2005; à la suite de van Inwagen 1999), on peut penser que les réflexions qui suivront dans cette section s'appliquent aussi bien à l'action qu'à l'omission (voir Cyr 2016). Deuxièmement, il faut immédiatement admettre qu'il y a un sens indirect de la responsabilité morale (Pereboom 2014; M. McKenna 2016) pour lequel le PAP est immédiatement faux. Supposons que Samuel doive rendre *Le Petit Prince* à la bibliothèque. Conscient que la libraire croit que le PAP est vrai et trop paresseux pour rendre le livre, Samuel le brûle (samedi). Dimanche, il n'a donc plus la possibilité de rendre l'exemplaire emprunté. Si le PAP est vrai, alors il y a un sens dans lequel Sam n'est plus responsable de ne pas rendre le livre à la bibliothèque – il serait, en quelque sorte, absurde de blâmer Samuel chaque jour jusqu'à ce qu'il rapporte l'exemplaire du *Petit Prince*. Mais il y a aussi un sens, indirect, dans lequel Samuel est responsable de rendre le livre. Puisqu'il convoque visiblement un sens indirect de la responsabilité morale, ce genre d'exemple n'est pas en général jugé troublant pour les défenseurs du PAP, qui peuvent en restreindre l'application à la responsabilité directe.

Pourquoi souscrire au PAP ? Si le principe peut sembler immédiatement intuitif, on lui connaît aussi quelques défenses. David Widerker (2000) suggère que pour qu'il soit juste de tenir une personne moralement responsable ou de la blâmer pour une action mauvaise, il faut qu'il existe une réponse raisonnable à la question « qu'aurais-je dû faire à la place ? ». Cette connexion entre *devoir* et *blâme* (ou entre la responsabilité comme *devoir* et la responsabilité comme connexion entre l'action et l'agent, condition nécessaire au blâme) suggère que nous devons admettre le principe suivant :

Blâme implique obligation : Si un sujet est blâmable pour une action (ou s'il est responsable d'une action mauvaise), alors il a enfreint l'un de ses devoirs.

Or, si l'on y ajoute

Devoir implique pouvoir : Si un sujet *doit* poser une action, alors il *peut* la poser.

On obtient :

Blâme implique pouvoir : Si un sujet est blâmable pour une action (ou s'il est responsable d'une action mauvaise), alors il pouvait agir autrement (c'est-à-dire qu'il pouvait satisfaire son devoir).

Ce qui revient, à peu de mots près, au PAP. (Un argument similaire est proposé dans Copp 2008). Le principe *Blâme implique obligation* suggère que les règles de la moralité doivent être utiles. Il serait de la nature des normes de pouvoir être suivies. C'est évidemment une question qui dépasse le cadre de cette entrée.

ii. Interpréter les possibilités alternatives

Avant de passer aux contre-exemples possibles au PAP, il faut se poser la question de la façon dont on devrait l'interpréter. C'est une question plus directement liée au libre arbitre, mais il est important de l'aborder brièvement. Que signifie *avoir pu faire autrement* ? Immédiatement, deux lectures sont possibles : il peut s'agir d'une *capacité* ou de l'*existence de possibilités accessibles*. Ces deux lectures semblent liées, mais on peut clairement voir qu'ils se séparent si l'on considère les cas de capacités masquées : un pianiste est assis dans un avion. A-t-il la capacité de jouer du piano ? S'il l'a, il y a quelque chose qu'il n'a pas, c'est-à-dire l'opportunité d'en jouer. A-t-il la possibilité d'en jouer ? Non.

Comme il serait absurde de déclarer responsable un pianiste de ne pas avoir joué du piano dans un avion qui n'en contient pas (on peut imaginer qu'un terroriste qui a détourné l'avion accepte d'interrompre son acte de terrorisme si le pianiste en question lui joue *Clair de Lune*), il apparaît *a priori* inconcevable de ne concevoir le contrôle que comme une notion liée à la capacité et non à l'opportunité (Hart 2008, 152).

Une fois cela admis, c'est-à-dire une fois que l'on tient compte d'un sens de « avoir pu faire autrement » qui inclut aussi bien la capacité que l'opportunité, nous pouvons encore diviser l'analyse en deux branches. La première est incompatibiliste, c'est-à-dire qu'elle considère que pour pouvoir faire autrement, il faut que le déterminisme soit faux (voir par exemple Kant 1781; Ginet 1966, 1990, van Inwagen 1983, 2000, 2008). En effet, pour qu'un agent puisse faire une action φ , il faut qu'il existe des mondes possibles dont le passé et les lois de la nature sont identiques aux nôtres, mais où l'action de l'agent n'est pas φ . Autrement dit, les possibilités qui sont accessibles sont en effet celles qui apparaissent dans des mondes dont les lois de la nature et l'état antérieur sont identiques ; mais si le déterminisme est vrai, le seul monde qui remplit ces conditions est le nôtre. La seconde est compatibiliste. Sa version classique, élaborée par David Hume (1748) consiste à dire qu'un agent peut faire une action (ou est libre de la faire) lorsque, *s'il avait voulu la poser*, il l'aurait fait. Les compatibilistes s'opposent donc à l'exigence incompatibiliste d'un monde possible où le futur est différent sans que le passé ou les lois de la nature le soient. Pour certains compatibilistes (p. ex. Lewis 1981; Vihvelin 2004, 2013; Taylor et Dennett 2011), il faudrait plutôt, par exemple, qu'il existe des mondes possibles dans lequel l'agent agit différemment et dont les lois de la nature ou le passé sont *suffisamment similaires* aux nôtres. Les débats sur le PAP concernent donc aussi bien les compatibilistes que les incompatibilistes.

iii. Les contre-exemples à la Frankfurt

Dans son influent article « *Alternate Possibilities and Moral Responsibility* », Harry G. Frankfurt propose un contre-exemple au PAP (Frankfurt 1969). En voici une paraphrase :

Contre-exemple à la Frankfurt

Alexandre voudrait que Sarah commette un vol de banque. Il sait qu'elle en a l'intention, mais pour s'assurer qu'elle le fasse, il place une puce dans son cerveau. Si Sarah décide de ne pas procéder au vol de banque, la puce s'allumera et la forcera à agir comme Alexandre le souhaite. Finalement, Sarah décide d'elle-même de voler la banque et la puce n'est jamais activée.

Frankfurt juge que Sarah est responsable de son action (et donc *prima facie* blâmable puisqu'il s'agit d'un vol qualifié), mais qu'elle n'aurait pas pu faire autrement. En effet, sa décision vient d'elle ; elle en est la *source*. L'agent contrefactuel Alexandre n'a au final pas eu de rôle causal sur la décision de Sarah ; elle en est donc entièrement responsable. Bien sûr, si elle avait voulu faire autrement, le dispositif aurait été activé. Peu importe ce qu'elle décidait, donc, elle finissait par commettre un vol de banque. En conséquence, nous n'aurions pas besoin de possibilités alternatives pour être responsables de nos actions. C'est un résultat important, qui a donné lieu à une littérature foisonnante, ne serait-ce que parce que le PAP est traditionnellement utilisé pour défendre l'incompatibilité entre le déterminisme et la responsabilité morale.

Les contre-exemples à la Frankfurt radicalisent une expérience de pensée de John Locke, en lui donnant une morale différente :

Supposons qu'un homme, alors qu'il est assoupi, soit transporté dans une pièce où se trouve une personne avec laquelle il désire s'entretenir depuis longtemps, et qu'il soit enfermé à l'intérieur, sans possibilité de sortir. Au réveil, il est heureux de se trouver en si bonne compagnie, en laquelle il demeure son plein gré [...]. (Locke 1836, chap. 21, section 10, traduction libre)

La morale que Locke en tire est la suivante : comme l'individu reste *volontairement* dans la pièce, mais qu'il n'est pas libre d'en partir et, croit Locke, par le fait même d'y rester, la liberté n'a rien à avoir avec la volonté (avec la « volition », dit-il). L'homme n'est pas libre, mais il agit volontairement. Au contraire, quelqu'un comme Frankfurt dirait : puisque l'individu reste volontairement dans la pièce, il est libre d'y rester ; et comme il n'est pas libre d'en partir c'est-à-dire qu'il n'a pas la possibilité de partir, l'on peut agir librement même faute d'alternative. Frankfurt ajoute : c'est ce genre de liberté qui assoit la responsabilité morale ; l'homme est donc responsable de rester dans la pièce. Bref, il serait possible de faire l'inévitable librement et d'en être responsable si cela est fait « de notre plein gré » ou « volontairement ».

La leçon des scénarios de Locke n'est-elle pas suffisamment explicite ? Pourquoi alors s'empêtrer dans des scénarios compliqués où des agents contrefactuels insèrent des dispositifs de contrôle ? La différence cruciale entre le scénario de Locke et celui de Frankfurt est qu'il apparaît assez clair pour certains que dans le scénario de Locke, l'homme n'est responsable que de ne pas avoir *tenté* de quitter la chambre et non de ne pas *l'avoir quittée*. Dans le cas de Frankfurt, comme le verrou de la porte est, pour ainsi dire, déplacé dans la tête de l'agent, il est plus difficile de formuler une réponse analogue.

iv. Trois réponses aux contre-exemples à la Frankfurt

Les réponses aux contre-exemples à la Frankfurt peuvent se décliner en trois catégories.

I. La première catégorie de réaction aux contre-exemples à la Frankfurt consiste en un rejet du contre-exemple à la Frankfurt et de certaines de ses variantes. C'est une réponse similaire, mais plus générale, à celle présentée ci-dessus pour répondre aux cas de Locke. La voici : si Sarah est responsable d'une action dans le cas de Frankfurt, ce doit être une action qu'elle aurait pu éviter. Il y a alors plusieurs choix : Sarah peut être responsable d'avoir effectué l'action *à ce temps-ci plutôt qu'à ce temps-là* (Ginet 1996) ; elle peut également être responsable de ne pas avoir *essayé de décider* (van Inwagen 1999; Widerker 2000, 2006; Moya 2011). Est-ce que ces possibilités sont toujours ouvertes ? C'est ce qu'essaient de démontrer les tenants de la « défense du dilemme » (Widerker 1995; Ginet 1996; Kane 1996; Goetz 2005; Palmer 2014) :

La défense du dilemme

Soit le déterminisme est vrai dans les cas de Frankfurt, soit il ne l'est pas.

- - Si le déterminisme est vrai, alors l'action de Sarah est déjà inévitable, du point de vue des incompatibilistes qui endossent le PAP. Le contre-exemple à la Frankfurt déterministe est donc entièrement impertinent pour eux, mais aussi pour tous ceux qui sont agnostiques sur la question (puisque s'ils sont agnostiques, c'est qu'ils ne considèrent pas que les cas ordinaires d'actions qui semblent être blâmables dans un monde déterministe permettent de trancher la question ; y ajouter un agent contrefactuel ne devrait changer à l'affaire).
 - Si le déterminisme est faux, alors Sarah aura toujours, peu importe la construction du cas de Frankfurt, la possibilité de faire autrement. Or, un scénario où Sarah a la possibilité de faire autrement ne pourra pas être un contre-exemple au principe qui postule que Sarah doit pouvoir faire autrement pour être moralement responsable.

Les réponses qui s'ensuivent sont nombreuses et complexes (Fischer 2010, 318, compare cette littérature à celle sur les cas de Gettier). Nous nous contenterons d'esquisser deux répliques possibles. La première consiste à attaquer la branche indéterministe de la défense du dilemme en reconstruisant un nouveau type de contre-exemple à la Frankfurt (Pereboom 2001, 2003, 2009, 2012, 2014; Hunt 2007). Voici un amendement qu'on peut joindre au scénario de Frankfurt présenté plus haut :

Amendement au contre-exemple à la Frankfurt

Alexandre sait que lorsque Sarah s'apprête à décider de faire la bonne chose moralement (ou même peut-être à initier ce genre de décision), elle rougit légèrement (elle ignore totalement qu'elle présente cette particularité physiologique). Alexandre programme son dispositif de contrôle de sorte qu'il se déclenchera si Sarah rougit : le cas échéant, Sarah n'initierait même pas sa décision de ne pas commettre un vol de banque. Le dispositif n'est jamais activé et Sarah commet un vol de banque.

L'idée de ce genre d'amendement est le suivant : l'absence de déterminisme ne suffira pas à donner des possibilités alternatives *robustes*, c'est-à-dire à la fois accessibles à l'agent et raisonnables. Le PAP est donc peut-être vrai à strictement parler, mais s'il l'est, c'est trivialement puisque même des possibilités alternatives non robustes seraient suffisantes à le satisfaire. Une version du PAP qui exigerait des possibilités alternatives robustes pour la responsabilité morale (ce qui est l'esprit véritable du PAP) serait invalidée par le contre-exemple amendé. (Pour une discussion détaillée des réponses subséquentes, voir (Pereboom 2014)).

Une deuxième réplique consiste à attaquer la branche déterministe de la défense du dilemme. D'abord, McKenna et Haji (2006) ont argumenté que la branche déterministe n'était pas problématique puisqu'il serait permis de ne pas supposer que le déterminisme supprime les possibilités alternatives. Cela est juste dans le sens où, comme nous l'avons déjà évoqué, les contre-exemples à la Frankfurt sont également importants pour les compatibilistes classiques. La question fondamentale est de déterminer si les contre-exemples à la Frankfurt ont une portée pour les parties « neutres ». Fischer suggère, dans la même veine, un argument en deux étapes (Fischer 2010, 2013). D'abord, la présence de l'intervenant contrefactuel suggère que le simple fait de ne pas avoir d'alternatives n'est pas pertinent pour l'attribution de la responsabilité morale. Ensuite, puisque le déterminisme ne menace la responsabilité morale qu'en ce qu'il supprime les possibilités alternatives, on constate que le déterminisme ne menace pas la responsabilité morale. À cela, plusieurs (Goetz 2005; Palmer 2005, 2014; Widerker et Goetz 2013) ont répondu que sans la présomption que le déterminisme supprime les possibilités alternatives, il est impossible de montrer comment la présence de l'agent contrefactuel mine ces mêmes possibilités alternatives.

II. Déterminer si les scénarios présentés par Frankfurt sont réellement des contre-exemples au PAP n'est peut-être pas crucial : ils sont suffisants pour nous donner une nouvelle perspective sur la responsabilité morale. La deuxième réaction aux contre-exemples à la Frankfurt est donc d'abord l'admission que le PAP est faux, mais ensuite et surtout que l'ingrédient central de la responsabilité morale est l'expression de notre *moi réel* (Frankfurt 1971; Watson 1975; D. W. Shoemaker 2011; Sripada 2015, 2016). Il y a donc élimination ou bien réinterprétation assez radicale de la notion de contrôle comme condition nécessaire à la responsabilité morale (les théoriciens du moi réel sont parfois des attributionnistes). On appelle ce genre de théorie « *mesh theories* » (« *mesh* » signifie « concorde », « engrenage », « filet ») ou « théories de l'identification ». Une action libre est une action qui a une forme de correspondance avec la personne.

Pour bien comprendre les théories de l'identification ou du moi réel, il faut garder en tête le compatibilisme classique de Hume qui offre une sorte de canevas aux théories de l'identification. De plus, les théories de l'identification permettent de résoudre le problème principal du compatibilisme classique de Hume. Pour Hume, un agent est libre de faire une certaine action lorsque, *s'il avait voulu poser cette action*, il l'aurait posée. On peut également dire la même chose de la responsabilité morale : un agent est responsable d'une action s'il était libre de faire autrement dans le sens où, *s'il avait voulu agir autrement*, il l'aurait fait. Le problème du compatibilisme classique est qu'il donne un diagnostic déconcertant pour les cas comme le suivant :

Héroïnomane simple : un héroïnomane veut consommer de l'héroïne et le fait sur une base quotidienne. Il a le désir de ne pas consommer d'héroïne, mais ce désir n'est pas

effectif. Il n'a pas le désir de ne plus désirer de l'héroïne et ne pourrait pas avoir ce type de désir, à savoir des désirs à propos de ses désirs.

Pour le compatibiliste classique, l'héroïnomane est libre d'arrêter de consommer, et ce, même s'il est psychologiquement impossible pour lui de le vouloir. En effet, *s'il avait voulu* arrêter de consommer, il l'aurait fait. C'est le même genre de constat que la théorie doit donner pour les cas où un agent est affligé par un trouble mental ou une phobie. Le problème du compatibilisme classique serait également celui des tenants du PAP en général, c'est-à-dire qu'il mettrait trop l'accent sur des faits modaux comme la capacité à faire autrement et pas assez sur la structure actuelle de l'agent.

Harry Frankfurt propose une théorie de l'identification (Frankfurt 1971), désormais classique, qui permet de pallier ce problème (Frankfurt (2002) a proposé d'autres modèles par la suite). Elle se compose essentiellement de *quatre distinctions*. D'abord, on dit qu'un agent *veut* faire une action aussi bien quand on parle de simples désirs *non-effectifs* que lorsqu'on parle de désirs *effectifs*, qui expliquent l'action (ou la causent) (Frankfurt 1971, 8). Notre héroïnomane simple a un désir effectif de prendre de l'héroïne et un désir non-effectif de ne pas en prendre.

Une deuxième distinction se situe entre deux ordres de désirs. Les désirs *de premier ordre* sont des désirs de poser certaines actions ; les désirs *de second ordre* sont des désirs d'avoir certains désirs de premier ordre. Habituellement, le désir de second ordre est un désir qu'un désir de premier ordre soit effectif : ainsi, j'ai un désir que mon désir de travailler soit effectif. C'est un *désir à propos de ma volonté*, ce que Frankfurt appelle une *volition de deuxième ordre*. Il y a des cas où le désir de second ordre est simplement d'avoir un désir, sans qu'il soit effectif, comme lorsqu'un thérapeute veut pouvoir comprendre le désir violent de son patient sans pour autant que ce désir violent soit effectif.

Troisièmement, Frankfurt distingue les *personnes* des autres créatures. Son critère principal est la possession de volitions de second ordre. Une personne a le désir d'avoir certains désirs opérants ; elle a un souci pour sa volonté (Frankfurt 1971, 12). L'héroïnomane simple n'est donc pas une personne puisqu'il n'a que des désirs de premier ordre. *A contrario*, l'héroïnomane réticent est une personne :

Héroïnomane réticent : Comme l'héroïnomane simple, l'héroïnomane réticent a un désir de premier ordre de consommer de l'héroïne et un désir de premier ordre de ne pas en consommer. Mais il a également une volition de second ordre de ne pas consommer d'héroïne.

L'héroïnomane réticent *s'identifie* à certains de ses désirs de premier ordre ; il les fait *siens*. C'est par cette identification qu'il donne un sens à l'affirmation selon laquelle ce n'est pas de son plein gré qu'il s'abandonne à l'héroïne (Frankfurt 1971, 13). Quant à l'héroïnomane simple, il ne s'identifie à aucun désir ; il ne se préoccupe pas du désir qui prendra le dessus.

Frankfurt trace une quatrième distinction, entre *agir avec une volonté libre ou un libre arbitre* (« acting with a will that is free ») et *agir de son plein gré* (« of one's own free will ») (voir M. McKenna et Coates 2015). Une personne *agit de son plein gré* est une personne dont les désirs effectifs de premier ordre sont en harmonie avec ses volitions de

deuxième ordre. L'héroïnomane réticent n'agit donc pas de son plein gré lorsqu'il consomme de l'héroïne. Par contre, l'héroïnomane assumé agit de son plein gré :

Héroïnomane assumé : Un héroïnomane a un désir de premier ordre de consommer de l'héroïne. Il a également un désir de second ordre de désirer consommer de l'héroïne (autrement dit, il aime être héroïnomane). Il ne peut pas se débarrasser de son désir de second ordre : par exemple, si sa dépendance venait à s'estomper, il tenterait de la ranimer.

Contrairement à l'héroïnomane réticent, l'héroïnomane assumé s'identifie à son désir de consommer de l'héroïne. Sa consommation est donc libre. Par contre, il y a bien quelque chose que l'héroïnomane assumé n'a pas et c'est un *libre arbitre*. Une personne a un *libre arbitre* lorsqu'elle est libre de vouloir ce qu'elle veut vouloir. Cela signifie qu'il est en son pouvoir de s'identifier à un désir de premier ordre ou à un autre : « tout comme la question de la liberté d'agir pour un agent dépend de si l'action est voulue [conformément au compatibilisme classique], la question de la libre volonté (libre arbitre) dépend de si la volonté est voulue. » (Frankfurt 1971, 15).

Frankfurt croit qu'un sujet est *moralement responsable* des actions qu'il effectue de son plein gré, c'est-à-dire qui sont conformes aux désirs de deuxième degré. Cela est vrai, que le sujet ait un libre arbitre ou non (Frankfurt 1971, 19). Puisque c'est cette conformité qui détermine ce qu'est la *volonté* d'un individu, Frankfurt réaffirme la doctrine selon laquelle nous sommes responsables des actes issus de nos volontés. Ainsi, l'héroïnomane simple n'est pas responsable puisqu'il n'a pas de volonté ; l'héroïnomane réticent n'est pas responsable puisque sa consommation va à l'encontre de sa volonté ; enfin, l'héroïnomane assumé est responsable puisque sa consommation est issue de sa volonté – il s'identifie à son désir de consommer.

Le fait que l'identification soit cruciale pour la responsabilité morale fait de la théorie de Frankfurt une théorie de la responsabilité morale de l'identification ou du « moi réel » (Wolf 1993, 29). Les actions dont un agent est responsable sont celles qui révèlent qui il est ; ce sont ses actions.

Gary Watson juge que le problème avec la théorie hiérarchique est que, tout comme il est possible de concevoir que nos désirs de premier ordre nous soient étrangers, il se peut que nos désirs de second ordre le soient aussi. Cela n'est pas tant motivé par l'idée qu'il faudrait remonter plus haut dans les ordres de désirs que par celle que les *désirs* eux-mêmes n'ont pas l'autorité suffisante pour déterminer s'il appartient à l'intérieur ou à l'extérieur de l'agent (Watson 1975, 1987, 149). Pourquoi un désir de plus haut ordre aurait-il plus d'autorité ? Certainement pas parce qu'il suppose plus de désirs (un désir de second et de premier ordre) (Wolf 1993, 29-30).

Contra Hume et Frankfurt, Watson est convaincu qu'évaluer ne se réduit pas à désirer et que l'évaluation a une autorité spéciale que le désir ne possède pas. Seules les évaluations donneraient des *raisons* de s'opposer à des désirs (et non simplement d'autres désirs). Watson propose donc de faire jouer le rôle de la distinction entre désirs de premier ordre et de second ordre à un nouveau couple : celui du système motivationnel et du système évaluatif (Watson 1975). Le *système motivationnel* est l'ensemble des attitudes non rationnelles comme les désirs, les impulsions et les émotions ; le *système évaluatif* est l'ensemble des attitudes rationnelles, qui, lorsqu'elles

sont combinées aux jugements factuels, donnent des jugements de valeur, d'obligation, etc. Dans le système de Watson, les jugements évaluatifs (ou simplement les valeurs) jouent le rôle des désirs de second ordre chez Frankfurt puisqu'elles (ou du moins une partie d'entre elles) constituent le moi réel d'un sujet. Ces valeurs sont peut-être des types de désir, mais si elles le sont, elles ont une particularité. Contrairement aux autres désirs, les valeurs sont telles que les agents ne voudraient pas y renoncer, et ce, même si ces désirs ne pouvaient être satisfaits. Par exemple, en réalisant qu'il ne peut réaliser la paix dans le monde, Gandhi ne devrait pas être amené à réviser ses valeurs. À l'inverse, certains de nos désirs qui ne sont pas des valeurs sont tels que nous souhaiterions parfois y renoncer plutôt qu'ils soient satisfaits d'une autre manière (Wolf 1993, 31). Si ce genre de théorie apparaît démesurément intellectualiste, il faut tenir compte de ses variantes, où, par exemple, l'accent est mis non sur les attitudes cognitives, mais sur les attitudes conatives comme le soin ou la préoccupation (Watson 1987, 150; Sripada 2016).

Malgré leur attrait certain, les théories du moi réel ou de l'identification semblent excuser à la fois trop et trop peu. Elles semblent d'abord *excuser trop* de types d'actions ou d'omissions. En voici trois. Il y a d'abord les actions ou omissions d'individus qui ont une structure évaluative déficiente, mais sans qu'ils soient incapables d'en former une. Dans *l'Étranger* de Camus, Meursault n'endosse pas son désir de tuer ; il semble être presque dénué de désirs de second ordre ou de valeurs. Cela ne semble pourtant pas le rendre moins responsable ; il était raisonnable d'exiger qu'il valorise certaines choses ou ait entretenu certains désirs puisqu'il a certainement la capacité d'entretenir des désirs de second degré ou de former des valeurs (il ne souffre d'aucune maladie ou de trouble). Un deuxième type de cas est celui de l'akrasie (Wolf 1993, 32). Par exemple, ceux qui se rendent coupables de crimes passionnels n'ont pas à s'identifier à leurs désirs violents ; au contraire, ils peuvent détester être jaloux et pourtant s'abandonner à leurs impulsions. Ils peuvent désirer ne pas désirer tuer ou encore ne pas avoir de valeurs violentes. Cela les rend-il automatiquement non responsables ? Cela semble difficile à admettre. Un troisième cas concerne la disparité qu'il y a parfois entre ce qu'un agent endosse rationnellement et les émotions profondes qu'il ressent. L'endossement rationnel est alors presque superficiel comparativement aux émotions qui, elles, manifestent le moi réel ou profond. Huckleberry Fin croit qu'il doit rendre un esclave à son maître et pourtant il est incapable de ne pas mentir à ses poursuivants afin de sauver l'esclave. Malgré tout, il a des *croyances morales* racistes et esclavagistes. Ne manifeste-t-il pas qu'*au fond* (de son moi), il est bon et que ses croyances racistes sont pour ainsi dire moins centrales que ses émotions humanistes ? Il n'est pas absurde de supposer même qu'il mérite d'être louangé (Arpaly et Schroeder 1999).

Le théoricien de l'identification peut régler ce dernier problème de deux façons : ou bien en donnant aux attitudes conatives (ou émotions) un rôle plus central, ou bien en donnant aux raisons objectives (plutôt que subjectives comme les valeurs) une certaine importance. Malgré tout, même s'il peut régler ce problème, le théoricien de l'identification se cogne aux deux premiers défis. Il semble que la seule façon de les relever est de supposer que ce qui fonde la responsabilité morale n'est pas tant la correspondance entre une action et certaines valeurs, mais une action et certaines *capacités* à avoir certaines valeurs.

Les théoriciens du moi réel semblent également *excuser trop peu*. Reprenons le cas de l'héroïnomane réticent, que les théoriciens du moi réel jugent non-responsable. Imaginons maintenant qu'il ait perdu la bataille et qu'il se soit complètement abandonné à l'héroïne. À un certain point, il en vient à oublier sa famille et donc à ne plus détester l'impact de l'héroïne sur sa vie. Ultimement, il endosse son désir pour la drogue ; il attache de la valeur à sa seule consommation. Il devient alors un héroïnomane assumé. Devient-il alors (comme par magie) moralement responsable de ses actions, et ce, même s'il semble simplement *plus profondément enfoncé* dans une forme d'auto-esclavagisme que l'héroïnomane réticent ?

Une façon de régler ce problème est de considérer que les actions libres et dont un sujet est moralement responsable sont les actions rationnelles ou qui correspondent aux raisons. L'héroïnomane assumé n'agirait pas rationnellement et serait donc non responsable. Le problème évident de cette suggestion est qu'elle rend toute action qui ne correspond pas aux raisons, et donc possiblement toute action immorale, non blâmable. Dans sa « théorie de la Raison » ou du *moi profond* par opposition au *moi réel*, Susan Wolf nous propose une solution de ce type (Wolf 1993). Wolf utilise toujours une majuscule pour écrire « Raison », peut-être pour signifier une allégeance à une thèse sur l'objectivité des raisons ou simplement pour s'inscrire dans la tradition platonicienne selon laquelle il est impossible de vouloir le mal (voir *Protagoras* 352b-357a ; *Ménon* 77b-78a). R. Jay Wallace propose une théorie similaire, où la capacité de comprendre les raisons morales est le cœur de la responsabilité (Wallace 1994, 2002).

Ce qui compte pour la responsabilité morale n'est ni la capacité d'agir autrement (ce que Wolf appelle l'autonomie) ni l'identification à un moi réel, mais la capacité d'agir en conformité avec la raison (Wolf 1993, 68). Puisque c'est la possession d'une capacité qui rend responsable, la conclusion que nul n'est responsable d'actions immorales est évitée. Il reste que Wolf propose une solution asymétrique : les gens qui suivent la Raison, mais qui sont incapables de faire autrement sont responsables, mais ceux qui ne suivent pas la Raison et qui sont incapables de faire autrement ne le sont pas. Autrement dit, la théorie de la Raison ne requiert pas toujours la capacité de faire autrement, mais seulement la capacité de faire *autrement que la mauvaise chose*.

Les deux types de problèmes à la théorie de l'identification – qu'elle excuse les actions qui ne sont pas identifiées au sujet, mais que le sujet contrôlait (Meursault et le crime de passion) et qu'elle n'excuse pas les actions qui sont identifiées au sujet, mais qui manifestent l'irrationalité ou le manque de contrôle profond du sujet (l'héroïnomane assumé) – mènent à une solution commune qui pose un problème. Il s'agit de se tourner, comme le suggère Wolf, vers une certaine *capacité*, celle d'agir conformément à la Raison. Mais alors, qu'est-ce qui fonde la responsabilité morale : le fait qu'une action est en harmonie avec les désirs ou les valeurs d'une personne ou le fait que cette personne a la capacité de faire la bonne action (et donc autrement) ? Le cas échéant, l'on s'éloigne de la théorie de l'identification. Pire : on se retrouve avec le problème initial des contre-exemples à la Frankfurt, où le sujet n'a *pas* cette capacité (pour une analyse contraire, voir par exemple M. Smith 2003; Vihvelin 2004, 2013, 2015; Fara 2008).

III. La troisième réponse face au contre-exemple à la Frankfurt est l'admission que le PAP est faux, mais le maintien de la notion de contrôle par la sensibilité aux raisons comme épicycle de la responsabilité morale (Fischer 1994, 2007; Fischer et Ravizza

1998; Sartorio 2015, 2016a). Autrement dit, cette troisième réponse considère que les autres théories du moi réel ou de l'identification échouent à donner à la *sensibilité aux raisons morales* la place qui lui revient. La *sensibilité aux raisons* est la capacité à réagir aux raisons, laquelle ajoute le contrôle manquant dans la théorie de Frankfurt. Cette troisième solution ne suggère pas pour autant de revenir d'une certaine façon à la *capacité de faire autrement* comme Susan Wolf, puisque ces théories ont pour point de départ le rejet du PAP. (Une autre différence avec la théorie de Wolf est qu'alors que Wolf considère que les raisons morales sont des faits objectifs, Fischer et Ravizza tentent d'être plus œcuméniques en se contentant de contraintes de cohérence et d'intelligibilité aux mécanismes qui sont sensibles aux raisons (Voir D. Zimmerman 2002 pour une évaluation critique de ce formalisme)). Nous laisserons de côté cette différence dans la discussion qui suit.)

Sans explorer le détail des différentes variantes de cette réponse aux contre-exemples à la Frankfurt, il faut mentionner certaines distinctions importantes sur lesquelles elles se fondent (voir M. McKenna 2006 pour un excellent résumé) et qui sont héritées du débat sur ces contre-exemples. C'est que, croient plusieurs intervenants dans le débat (Frankfurt 2003; Fischer 2010; Sartorio 2015), l'important des scénarios à la Frankfurt n'est pas qu'ils constituent de réels contre-exemples au PAP.

Les contre-exemples à la Frankfurt permettent d'abord de distinguer clairement deux concepts : le fait de rendre inévitable et le fait de faire advenir une action (Frankfurt 2003, 339-40). Cela permet de distinguer deux types de faits dans une situation : ceux qui ont à voir avec la *séquence actuelle* comme la causalité et ceux qui ont à voir avec les *propriétés modales* de la situation comme les capacités (il faut mettre de côté ici l'idée selon laquelle la causalité pourrait être entièrement réductible à des faits contrefactuels). Ainsi, la troisième réponse aux contre-exemples à la Frankfurt met l'accent sur la séquence actuelle (on parle de « *actual sequence model* ») plutôt que sur les propriétés modales. Pourquoi mettre l'accent sur la séquence actuelle ?

L'importance de la séquence actuelle peut d'abord être montrée par la négative. Les contre-exemples à la Frankfurt rappellent la priorité de la séquence actuelle sur les faits modaux parce que notre jugement de responsabilité montre que les agents contrefactuels qui n'interviennent pas ne sont pas importants, et ce, même s'ils sont essentiels pour les faits modaux comme la capacité de faire autrement. (*A contrario*, les partisans de la théorie dispositionnelle du libre arbitre (p. ex. Vihvelin 2004; Fara 2008) croient que les contre-exemples à la Frankfurt préservent les faits modaux importants pour la responsabilité morale, c'est-à-dire la capacité de faire autrement). Carolina Sartorio croit même que les éléments qui ne font pas partie de l'explication causale d'une action (dont la présence d'agents contrefactuels) sont tout simplement hors de propos dans la détermination de la responsabilité morale (Sartorio 2015).

Par la positive, les contre-exemples à la Frankfurt illustrent que le contrôle se décline sous plusieurs formes : l'un dépendrait surtout de la séquence actuelle et l'autre des faits modaux. C'est à John Martin Fischer que nous devons cette distinction entre deux types de contrôle, à savoir le contrôle de guidage (*guidance control*) et le contrôle de régulation (*regulative control*) (Fischer 1994, 131-59). Le *contrôle de guidage* n'implique que la liberté pour un agent de « guider » sa conduite. Le *contrôle de régulation* suppose quant à lui quelque chose que l'agent dans les contre-exemples à la Frankfurt n'a pas, c'est-à-dire la possibilité de faire autrement. Or, puisque l'agent est moralement

responsable dans les contre-exemples à la Frankfurt, cela signifie que c'est seulement le contrôle de guidage et le contrôle de régulation qui est nécessaire à la responsabilité morale. (C'est pourquoi Fischer est un « semi-compatibiliste » : il croit que le déterminisme est incompatible avec le libre arbitre et avec le contrôle de régulation, mais pas avec la responsabilité morale.)

Qu'est-ce que signifie pour un agent de *guider* sa conduite ? Fischer croit que le mécanisme qui cause la conduite doit être *sensible aux raisons*, c'est-à-dire qu'il doit y avoir des situations dans lesquelles si une raison différente était accessible, le mécanisme répondrait différemment. Cela revient-il à postuler l'existence de la capacité à agir autrement ? Pas exactement : pour le montrer, il faut reprendre un cas à la Frankfurt. Sarah commet un vol de banque après l'avoir décidé. Il y a une puce dans son cerveau qui lui enlève la capacité de faire autrement. Le mécanisme qui produit sa décision est-il sensible aux raisons ?

Pour donner une réponse, il faut déterminer le niveau de sensibilité aux raisons qui est nécessaire. Les partisans de la troisième réponse comme John Martin Fischer et Mark Ravizza croient que la sensibilité aux raisons suffisante pour la responsabilité morale est une sensibilité aux raisons modérée. La sensibilité aux raisons modérée est caractérisée par trois éléments (Fischer et Ravizza 1998, 82, 242-43) : d'abord, le mécanisme est régulièrement réceptif aux raisons, dont certaines sont morales ; ensuite, il y a un scénario possible proche dans lequel le mécanisme aurait répondu à une raison ; enfin, le mécanisme est propre à l'agent (nous reviendrons plus tard sur ce dernier point).

Prenons un cas de Frankfurt pour tester la théorie. Sarah s'apprête à braquer une banque après avoir délibéré ; une puce a été implantée dans son cerveau, mais elle n'a pas été activée. Sarah est-elle responsable ? Le mécanisme qui a produit son action est, disons, son cerveau, ce qui comprend le système de valeurs de Sarah, son raisonnement, etc. Ce mécanisme est régulièrement réceptif aux raisons morales, puisque Sarah n'est pas une personne complètement immorale à l'habitude. C'est aussi le mécanisme *de Sarah* ; c'est un mécanisme qu'elle s'est approprié depuis longtemps. Est-ce qu'il y a un scénario possible proche dans lequel le mécanisme aurait répondu à une raison ? On pourrait initialement en douter : après tout, si par exemple, Sarah était informée de la présence d'un corps policier aux abords de la banque, la puce serait activée par Alexandre et Sarah aurait quand tenterait de braquer la banque. En revanche, nous ne nous posons pas la question si *Sarah* est sensible aux raisons, mais si *le mécanisme* qui produit son action l'est. Pour déterminer si ce mécanisme est sensible aux raisons, le monde dans lequel Sarah est informée de la présence de la police et décide de ne pas braquer la banque et où la puce n'est pas dans la tête de Sarah est suffisamment proche.

On peut penser que Fischer et Ravizza se facilitent la tâche en n'incluant pas la puce dans le mécanisme pertinent à l'évaluation de la responsabilité morale de Sarah. En effet, en incluant ce mécanisme, le deuxième critère ne serait pas satisfait : si Sarah était informée de la présence de la police, le mécanisme produirait la même décision, soit celle de braquer la banque. Ce qui permet à Fischer et Ravizza de ne pas inclure la puce dans le mécanisme est le fait que la puce ne contribue pas à *causer* l'action dont l'agent est présumé responsable. Leur théorie demeure en ce sens résolument volitionniste au sens classique du terme : un agent est responsable d'une action lorsque cette action est *causée de la bonne manière*. Comme la causalité n'est pas supprimée par l'absence de certaines propriétés modales comme la capacité de faire autrement, Fischer et Ravizza

n'ont pas à la postuler. Il reste qu'il est difficile de ne pas donner d'argument circulaire en faveur de leur analyse. L'individuation des mécanismes demeure donc un problème pour la théorie de Fischer et de Ravizza. En effet, qu'est-ce qui empêche de considérer que le mécanisme pertinent est l'agent lui-même ? Cela nous ramènerait encore à la capacité d'agir autrement (Ginet 2006; M. McKenna 2013).

Il faut noter toutefois que l'idée de réinterpréter la condition de contrôle nécessaire à la responsabilité morale sans pour autant adopter une théorie de l'identification ne passe pas forcément par une théorie de la séquence actuelle. La théorie dispositionnelle du libre arbitre (M. Smith 2003; Vihvelin 2004, 2013, 2015; Fara 2008) entend montrer que l'agent dans les contre-exemples à la Frankfurt dispose du contrôle nécessaire, mais que c'est parce qu'il maintient sa capacité de faire autrement, comprise comme disposition (voir M. McKenna et Coates 2015). C'est une théorie qui donne un verdict similaire à celui de Fischer et Ravizza, au point où certains croient qu'il n'y a qu'une dispute verbale entre les partisans de la théorie dispositionnelle et ceux de la théorie des mécanismes sensibles aux raisons (voir sur ce sujet Franklin 2015; M. McKenna et Coates 2015; Cyr 2017).

b. Manipulation et conditions historiques (et externalistes)

i. Les arguments de manipulation

Le problème de la réinterprétation de la capacité de contrôle nécessaire pour la responsabilité morale (et donc des deuxième et troisième réponses aux contre-exemples à la Frankfurt présentées dans la dernière section), c'est qu'elles peuvent prêter le flanc à la critique à travers des arguments dits de la « manipulation » (Mele 1995, 2009, Pereboom 2001, 2013b, 2014, Todd 2011, 2013, 2016). À proprement parler, les arguments de la manipulation visent à démontrer la vérité de l'incompatibilisme (l'incompatibilité entre le déterminisme et la responsabilité morale). Ils présentent une série de cas similaires (de deux à quatre), à commencer par un cas de manipulation directe et invasive, où il semble qu'entre chaque étape il n'y a pas de différence de responsabilité morale. Comme le dernier cas n'est qu'une situation où un agent agit dans un monde où le déterminisme est vrai, les tenants de l'argument de la manipulation concluent que c'est en fait le déterminisme (ou l'une de ses conséquences) qui sape la responsabilité morale dans chacun des cas.

Or, les arguments de la manipulation sont aussi utilisés pour mettre en étau les théories de la séquence actuelle et les théories de l'identification. Ils essaient de montrer que l'identification ou les propriétés de la séquence causale sont insuffisantes pour déterminer la responsabilité morale : il faudrait, en plus, que l'agent ait une *histoire* adéquate. L'idée fondamentale est que les scénarios de manipulation proposent des cas où un individu a la bonne « structure mentale » ; ses désirs sont harmonisés, il a le plein contrôle de ses actions, etc., mais qui a cette structure pour des raisons qui semblent miner sa responsabilité morale. Comme le souligne Gary Watson,

Pour le compatibilistes, les conditions constitutives de la libre agencité ne dépendent pas conceptuellement de ses origines. Dans ce sens, la libre responsable agencité n'est pas une notion historique. Conséquemment, le compatibilisme implique la possibilité conceptuelle qu'un agent libre et responsable ainsi que l'exercice libre et responsable de son agencité soient le produit de créateurs tout-puissants. [...] Cette possibilité

découle du point général que les conditions de la responsabilité ne dépendent pas nécessairement de leur origine causale. (Watson 1999, 360-61)

Il semble qu'en somme les arguments de la manipulation constituent une porte d'entrée idéale pour les discussions sur les possibles conditions historiques de la responsabilité morale.

D'abord et avant tout, il convient de présenter un exemple d'argument de la manipulation tiré de Derk Pereboom (2014, 74-80) et adapté pour les fins de la cause.

Argument de la manipulation

Contexte : Dans tous les cas suivants, le professeur Plum décide de tuer M. White pour satisfaire ses ambitions personnelles et réussit à le faire. L'action considérée est la décision d'exécuter White. Cette action remplit les critères compatibilistes (de contrôle) suivants : (1) sa décision est cohérente avec le caractère du personnage sans que son désir soit irrésistible (cf. Hume 1739) ; (2) le désir effectif de Plum de tuer White est conforme à ses désirs de second ordre (cf. Frankfurt 1971) ; (3) la décision de Plum est issue d'un mécanisme sensible aux raisons : s'il avait cru que les conséquences de son geste seraient mauvaises pour lui, il se serait abstenu (cf. Fischer et Ravizza 1998).

Manipulation : Une équipe de neurochirurgiens manipule les états neuronaux de Plum grâce à une technologie radio. Cette intervention permet de produire déterministiquement un raisonnement qui conduit Plum à décider d'assassiner White. Sans l'intervention des scientifiques, Plum n'aurait pas été assez égoïste pour tuer White.

Création : Comme dans **Manipulation**, excepté que les scientifiques ont programmé Plum d'avance de sorte qu'il a un tempérament très égoïste qui se manifeste à l'occasion. En particulier, les scientifiques ont programmé Plum de façon à ce qu'il prenne la décision de tuer White.

Dressage : Comme dans **Création**, excepté que les scientifiques sont remplacés par une société qui a fait développer des tendances égoïstes à Plum pendant son enfance par un processus de dressage méticuleux, avant qu'il puisse les réviser. Cela le détermine à tuer White comme dans les cas précédents.

Déterminisme : Comme dans **Dressage**, excepté qu'il n'y a pas de société qui a dressé Plum d'une façon particulière. Par contre, les conditions du monde avant la naissance de Plum jointes aux lois de la nature impliquent que Plum décide de tuer Plum et s'exécute.

Comme Plum semble clairement ne pas être responsable de son action dans **Manipulation** et qu'il n'y a aucune étape où il devient responsable, avance Pereboom, il faut conclure que Plum n'est pas responsable dans **Déterminisme**. Or, s'il ne l'est pas, c'est parce que le déterminisme est incompatible avec la responsabilité morale.

Il existe un deuxième type d'argument de la manipulation, plus succinct, qui commence directement par un épisode similaire à **Création** : un dieu, Diane, crée un zygote pour que l'homme qui en naîtra, Ernie, pose une action à l'âge de trente ans (en satisfaisant

les quatre conditions de contrôle précédemment nommées). La deuxième étape, similaire à **Déterminisme**, remplace Diane par les lois aveugles de la nature.

ii. Trois réponses aux arguments de manipulation

Comme chaque étape de l'argument de Pereboom satisfait les conditions de contrôle compatibilistes non historiques généralement admises, il faut alors adopter l'une des trois options suivantes.

I. La première consiste à donner raison à Pereboom : les conditions de contrôle compatibilistes seraient insuffisantes parce qu'il y aurait également des conditions de contrôles *incompatibilistes* à respecter pour la responsabilité morale. On pense ici à un contrôle de régulation (et donc une capacité de faire autrement incompatible avec le déterminisme) ou au fait d'être la source ultime de nos actions. Ceux qui acceptent les arguments de la manipulation sont donc ou bien libertariens (des incompatibilistes non sceptiques), ou bien sceptiques. Il y a plusieurs formes de scepticismes et de libertarianismes, dont on pourra retrouver un sommaire dans l'entrée sur le libre arbitre.

II. La deuxième option revient à postuler que les conditions de contrôle compatibilistes *non historiques* épuisent la notion de contrôle et que la manipulation n'entrave la responsabilité morale que lorsqu'elle élimine l'une ou l'autre de ces conditions (Watson 1999; Frankfurt 2002, 2003; M. McKenna 2012; Ebels-Duggan 2013). Si les conditions de contrôle sont effectivement réunies dans **Manipulation**, alors Plum y est responsable. Frankfurt est on ne peut plus clair :

Si une personne fait quelque chose parce qu'elle le veut et que ce désir est sans réserve, que cette personne soutient pleinement ce désir, alors – dans la mesure où sa responsabilité morale pour exécuter cette action est concernée – il n'importe pas du tout comment cette personne en est arrivée là. [...] Un manipulateur peut réussir, à travers son intervention, à conférer à une personne non pas seulement des pensées ou des sentiments particuliers, mais un nouveau caractère. Cette personne est moralement responsable des choix et de la conduite où ce caractère les mène. (Frankfurt 2002, 27-28).

Frankfurt est donc un partisan de l'internalisme et du structuralisme de la responsabilité morale :

L'internalisme de la responsabilité morale : La responsabilité morale d'un sujet *S* pour une action φ survient sur les propriétés internes de *S* durant la séquence pendant laquelle il effectue φ . Plus simplement, la responsabilité morale ne dépend pas de facteurs extérieurs au sujet au moment de l'action.

Le structuralisme (ou l'ahistoricisme) de la responsabilité morale : La responsabilité morale d'un sujet *S* pour une action φ survient sur les propriétés de la séquence pendant laquelle *S* effectue φ . Plus simplement, la responsabilité morale ne dépend pas de facteurs historiques (antérieurs à la séquence de l'action du sujet).

Ces deux positions s'opposent à :

L'externalisme de la responsabilité morale : l'internalisme de la responsabilité morale est faux.

L'historicisme de la responsabilité morale : le structuralisme de la responsabilité morale est faux.

L'internalisme implique donc le structuralisme et l'historicisme implique l'externalisme. Il y a d'autres théories que celle de Frankfurt qui sont conformes à l'internalisme et au structuralisme : par exemple, celles qui mettent l'accent sur la capacité de faire autrement comme condition nécessaire et (presque) suffisante à la responsabilité morale.

Il faut noter que le structuralisme laisse place à l'histoire de l'action de trois manières. Premièrement, le structuralisme ne suppose pas que les considérations historiques sont *impertinentes* ; elles sont, après tout, souvent utiles pour déterminer les propriétés actuelles de l'agent, soit ses capacités (Vargas 2006, 356) ou ses attitudes (Ebels-Duggan 2013, 157-60). Seulement, le partisan du structuralisme ne croit pas qu'il soit *toujours nécessaire* de considérer les propriétés historiques d'un agent avant de le déclarer responsable. La question est de savoir si, par exemple, l'hypnose ou la manipulation ont laissé intactes les structures pertinentes à la responsabilité morale, comme certaines capacités. Deuxièmement, la séquence d'une action peut inclure la délibération qui y mène : la période pertinente pour le défenseur du structuralisme peut être assez longue. Troisièmement, le structuralisme accepte qu'un agent puisse être *indirectement responsable* d'une action ou d'une omission parce qu'il s'est précédemment placé dans une situation où il lui était raisonnable de croire que cette action ou omission deviendrait inévitable ou irrésistible (Vargas 2006; M. McKenna 2016, 86-87). Bien sûr, cette notion de responsabilité indirecte considère d'emblée des éléments antérieurs à la séquence de l'action. Un cas discuté fréquemment est celui de deux conducteurs complètement ivres, Sam et Jan, qui frappent un enfant. Sam est devenu ivre par sa propre décision et Jan a été forcé à boire et à conduire. La distinction entre la responsabilité directe et indirecte nous permet de dire qu'aucun n'est directement responsable, mais que Sam – et pas Jan – est indirectement responsable. Notons que la responsabilité indirecte n'est pas moins importante que la responsabilité directe.

L'internalisme et le structuralisme sont un couteau à double tranchant. D'un côté, lorsqu'ils sont combinés avec l'idée selon laquelle nous avons une capacité absolue de faire autrement ou que nous sommes la source ultime de nos actions, ils semblent atteindre les plus hauts standards de justice, puisqu'ils sont alors résolument incompatibles avec la *chance morale* (Voir Nagel 2012, pour une discussion de l'opposition entre justice et chance morale). L'internalisme implique que le conducteur en état d'ébriété qui ne frappe personne est tout aussi responsable que le conducteur en état d'ébriété qui frappe un enfant. D'un autre côté, lorsque le structuralisme est combiné avec la théorie d'identification de Frankfurt, on arrive à des verdicts qui nous apparaissent difficilement acceptables, par exemple sur la responsabilité morale des enfants-soldats devenus adultes. C'est pourquoi ce genre de stratégie est souvent désigné sous le nom de « ligne dure ». Certains suspectent toutefois que si le structuralisme ou l'internalisme peuvent sembler injustes, c'est uniquement parce que nous associons implicitement la responsabilité morale à la punition (Arpaly 2006, 36-37; Ebels-Duggan 2013, 147).

Quoiqu'il en soit, ce n'est pas forcément de bon cœur que l'internalisme est adopté (voir p. ex. Watson 1999). Il faut voir que Frankfurt et d'autres croient que c'est le prix du compatibilisme :

Nous sommes inévitablement façonnés et maintenus, après tout, par des circonstances sur lesquelles nous n'avons aucun contrôle. Les causes dont nous sommes l'objet peuvent alors nous changer radicalement, sans par-là faire en sorte que nous ne soyons pas des agents moralement responsables. Il est tout à fait égal que ces causes soient ou bien opérantes en vertu de forces naturelles qui modifient notre environnement ou bien qu'elles opèrent à travers le dessin de manipulateurs humains. (Frankfurt 2002, 28)

C'est en partie cette dernière thèse que remettent en question certains tenants de la troisième réponse aux arguments de la manipulation.

III. Cette troisième option consiste à maintenir encore une fois que les conditions de contrôle compatibilistes épuisent la notion de contrôle, mais d'ajouter des conditions supplémentaires à la responsabilité morale : des conditions *historiques*, ou, plus généralement, *externalistes*, lesquelles ne seraient pas satisfaites dans toutes les étapes des arguments de la manipulation. Les scénarios de manipulation (et pas forcément seulement les *arguments* de la manipulation) offrent, selon certains (p. ex. M. McKenna 2016, 84), la motivation principale pour ajouter des conditions historiques.

Voici quelques-unes de ces conditions historiques :

- **Une condition interpersonnelle** (e.g. Lycan 1987; Nozick 1997; Lehrer 2016). Cette première condition entraîne un rejet direct des arguments de la manipulation puisqu'il s'agit tout simplement de postuler que l'existence d'un manipulateur ou d'un intimidateur fait une différence. Autrement dit, il y aurait une condition *interpersonnelle* à la responsabilité morale. Le contrôle nécessaire à la responsabilité morale est donc en somme *moralisé* : il ne s'agit pas de capacités indépendantes des propriétés morales de la situation. Cela rappelle entre autres la thèse de Robert Nozick qui veut que la prison n'est pas une entrave à la liberté lorsqu'elle est moralement juste (Nozick 1997). Ainsi, ceux qui adoptent ce genre de condition à la responsabilité morale peuvent croire que l'existence de possibilités alternatives n'est pas nécessaire, mais que Sarah est responsable dans le contre-exemple à la Frankfurt présenté ci-dessus puisque *c'est un agent* (Alexandre) qui prive Sarah de ses possibilités alternatives (Lehrer 2016).

Cette position est plutôt isolée : il est assez commun, par exemple, de différencier la responsabilité morale des agents qui sont menacés ou intimidés en fonction de l'impact de la menace ou de l'intimidation sur leurs raisons d'agir (la menace ou l'intimidation elle-même ne changeant rien à l'affaire) (voir p. ex. Frankfurt 1969; Hyman 2013) ; il est également commun de discréditer l'idée selon laquelle un manipulateur fait la différence en proposant d'y substituer une maladie dégénérative ou un phénomène naturel incontrôlable (voir par exemple Pereboom 2001, 115, 2014, 79; Mele 2016, 77; Todd 2013).

- **Une condition doxastique d'appropriation de mécanismes** (Fischer 1994, 208-9, 2000, 389-90, Fischer et Ravizza 1998, 210, 242-43, 252, 2000). Pour endosser

ou s'approprier un mécanisme de production d'actions, un sujet doit acquérir certaines croyances à propos de lui-même, croyances qui peuvent être simplement dispositionnelles et donc non forcément conscientes (Fischer et Ravizza 1998, 214-18). Ce type de condition peut aussi bien déterminer si un agent est en général exclu de la sphère de la responsabilité (comme les enfants) ou si c'est une action qu'il exécute qui est soustraite de la sphère de la responsabilité (Eshleman 2001, 483-86).

Voici les trois conditions doxastiques (trois types de croyance qu'un sujet doit avoir) nécessaires à l'endossement d'un mécanisme. D'abord, un sujet doit acquérir la croyance, sur le long terme, que ses choix et actions (le type d'action ou de choix liés au mécanisme en question) ont en général une certaine efficacité causale, comme lorsqu'un enfant réalise qu'il a la capacité de faire mal à sa petite sœur. Ensuite, le sujet doit se voir, également pendant un certain temps, comme un participant du « jeu social de la responsabilité », qui implique l'indignation, le ressentiment et la gratitude. Ces deux croyances doivent être acquises sur la base d'états évidentiels ; elles ne peuvent pas apparaître par magie ou sous le coup de la manipulation. Cette distinction aide à tracer la distinction cruciale entre endoctrinement et éducation. Cela permet donc possiblement de régler le problème de la manipulation.

Si chacune de ces conditions comporte un lot de problème, il faut surtout souligner que la deuxième condition, qui postule qu'un agent doit se voir comme l'objet approprié d'attitudes réactives pour le type d'action en question, ne semble pas nécessaire. Andrew S. Eshleman suggère qu'un étudiant qui croirait à une incapacité (non pathologique) à se lever le matin, telle qu'elle l'excuserait de ne pas avoir assisté à un cours, pourrait tout de même en être responsable (Eshleman 2001, 481-88). Pour sa part, Alfred R. Mele remet en doute qu'un réel sceptique de la responsabilité morale s'en trouverait moins responsable (Mele 2006b). Plus généralement, l'idée que les conditions d'appropriation d'un mécanisme de décision ou d'action soient doxastiques semble suspecte (Eshleman 2001, 489-90).

- **Une condition historique sur les valeurs, le caractère ou les raisons morales de l'agent** (Wallace 1994; Mele 1995, 2006a, 2009, 2016; Haji 2009; M. McKenna 2016). La condition est parfois formulée positivement et parfois négativement. Positivement, la théorie peut, par exemple, affirmer que les valeurs inaltérables (« *unsheddable* ») cruciales à l'explication de l'action d'un agent ont pu être évaluées (l'agent en a eu l'opportunité et la capacité), révisées et admises par lui et ces évaluations, révisions et admission auraient été ou ont été le fait de l'exercice de capacités de cet agent (M. McKenna 2016, 97). Les valeurs sont dites « inaltérables » lorsqu'elles ne peuvent être altérées, ni en existence ni en degré, lors de contextes normaux de délibération. Ce genre de critère n'est pas doxastique : rien en principe n'exclut qu'un agent puisse croire qu'il n'a pas eu l'opportunité ou la capacité de réviser ses valeurs, mais qu'il soit responsable malgré tout parce qu'il fait erreur. De surcroît, cette condition historique ne dit rien des actions qui ne sont pas issues de valeurs inaltérables.

Cette condition est parfois formulée négativement. Par exemple, les capacités impliquées dans l'action doivent avoir été acquises sans compulsion (Mele 1995, 2006a, 2016). Une capacité acquise sans compulsion est une capacité acquise sans que les

capacités de contrôle d'un agent aient été contournées. Bien sûr, si ce contournement a été planifié volontairement par l'agent, il n'en perd pas sa responsabilité.

Deux difficultés peuvent se présenter face aux conditions historiques sur les valeurs. La première est une libéralité excessive dans l'admission des valeurs. S'il y a trop de valeurs – McKenna mentionne le fait de valoriser le yogourt grec –, alors il est trop facile qu'une action soit issue d'une valeur inaltérable qui n'est pas le résultat d'un processus adéquat. La deuxième difficulté est le piège de la régression à l'infini (Voir, par exemple, G. Strawson 1994; Barnes 2016). Supposons que *S* fait *A* sur la base de la valeur inaltérable *V*. *S* est responsable de *A* puisque *S* a eu l'opportunité d'amender *V* auparavant. Si *S* a eu l'opportunité d'amender *V*, alors *S* avait au moins une autre valeur *V2* à ce moment (sinon, sur quelle base amender *V*?). *A* a-t-il eu l'opportunité d'évaluer *V2* ? Si oui, les questions continuent, à l'infini. Sinon, n'est-il pas injuste de juger *S* sur la base d'une valeur qu'il a acquise à cause d'une valeur inaltérable qu'il n'a pas pu évaluer ? Cela dit, ce défi n'est pas une surprise pour une théorie compatibiliste de la responsabilité morale.

En endossant ces conditions historiques, les historicistes peuvent tracer une distinction entre des scénarios comme **Manipulation** et **Création** (Fischer 2016; Mele 2016). Dans **Création**, l'agent a une histoire où il a l'opportunité de ne pas adopter la ou les valeurs qui le conduisent à agir (même s'il est déterminé à ne pas saisir cette opportunité). Du même coup, les historicistes peuvent donner raison à notre tendance à adoucir notre blâme contre certains criminels endurcis lorsque nous en apprenons davantage sur leur enfance atroce (voir par exemple Watson 1996; M. McKenna 2016)..

4. Conclusion

La responsabilité morale est un domaine de recherche foisonnant, touchant la métaphysique, l'éthique, l'épistémologie et la métaphilosophie. La notion de responsabilité morale, nous l'avons montré, est en connexion étroite autant de notions que sont les obligations, le blâme et la louange, le libre arbitre, l'histoire, la manipulation, la causalité, etc. Bien d'autres sujets ont cependant été laissés pour compte : l'histoire de la notion de responsabilité sur la longue durée, le lien entre conscience et responsabilité, le débat sur la distinction entre la responsabilité morale et criminelle et la responsabilité collective.

Malgré tout, cette entrée a voulu donner une vision d'ensemble des débats sur la responsabilité et en particulier des différents concepts et conceptions de la responsabilité ; de ses liens avec les attitudes réactions ; enfin, des principales conditions nécessaires à la responsabilité morale.

5. Bibliographie

Adams, Robert Merrihew. 1985. « Involuntary Sins ». *The Philosophical Review* 94 (1): 3-31. doi:10.2307/2184713.

Arpaly, Nomy. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. New York ; Oxford: Oxford University Press.

———. 2006. *Merit, Meaning, and Human Bondage: An Essay on Free Will*. Princeton, N.J.: Princeton University Press.

Arpaly, Nomy, et Timothy Schroeder. 1999. « Praise, Blame and the Whole Self ». *Philosophical Studies* 93 (2): 161-88. doi:10.1023/A:1004222928272.

Assemblée nationale du Québec. 2016. « Responsabilité ministérielle ». *Encyclopédie du parlementarisme*.

Ayer, A.J. 1954. « Freedom and Necessity ». In *Philosophical Essays*. Oxford University Press.

Barnes, Eric Christian. 2016. « Character Control and Historical Moral Responsibility ». *Philosophical Studies* 173 (9): 2311-31. doi:10.1007/s11098-015-0610-2.

Braham, Matthew, et Martin van Hees. 2012. « An Anatomy of Moral Responsibility ». *Mind* 121 (483): 601-34. doi:10.1093/mind/fzs081.

Cane, Peter. 2016. « Role Responsibility ». *The Journal of Ethics* 20 (1-3): 279-98. doi:10.1007/s10892-016-9235-8.

Caruso, Gregg D. forthcoming. « Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model ». *Southwest Philosophy Review* 32 (1).

Caruso, Gregg D., et Stephen G. Morris. 2016. « Compatibilism and Retributivist Desert Moral Responsibility: On What Is of Central Philosophical and Practical Importance ». *Erkenntnis* 82 (4): 1-19. doi:10.1007/s10670-016-9846-2.

Clarke, Randolph K. 2005. « On an Argument for the Impossibility of Moral Responsibility ». *Midwest Studies in Philosophy* 29 (1): 13-24. doi:10.1111/j.1475-4975.2005.00103.x.

———. 2014. *Omissions: Agency, Metaphysics, and Responsibility*. New York, NY: Oxford University Press.

———. 2016. « Moral Responsibility, Guilt, and Retributivism ». *The Journal of Ethics* 20 (1-3): 121-37. doi:10.1007/s10892-016-9228-7.

Cogley, Zac. 2013. « The Three-Fold Significance of the Blaming Emotions ». In *Oxford Studies in Agency and Responsibility*, édité par David Shoemaker, 205–224. Oxford University Press.

Copp, David. 2008. « 'Ought' Implies 'can' and the Derivation of the Principle of Alternate Possibilities ». *Analysis* 68 (1): 67-75. doi:10.1093/analysis/68.1.67.

Cyr, Taylor W. 2016. « Moral Responsibility for Actions and Omissions: A New Challenge to the Asymmetry Thesis ». *Philosophical Studies*, décembre, 1-9. doi:10.1007/s11098-016-0851-8.

———. 2017. « Semicompatibilism: No Ability to Do Otherwise Required ». *Philosophical Explorations*.

Darwall, Stephen L. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, Mass. ; London: Harvard University Press.

- De Brigard, Felipe. 2015. « Consciousness and Moral Responsibility ». *Analysis* 75 (4): 661-67. doi:10.1093/analys/anv012.
- Dennett, Daniel C. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. MIT Press.
- Ebels-Duggan, Kyla. 2013. « Dealing with the Past: Responsibility and Personal History ». *Philosophical Studies* 164 (1): 141-61. doi:10.1007/s11098-013-0090-1.
- Eshleman, A. S. 2001. « Being is not Believing: Fischer and Ravizza on Taking Responsibility ». *Australasian Journal of Philosophy* 79 (4): 479-90. doi:10.1080/713659304.
- Fara, Michael. 2008. « Masked Abilities and Compatibilism ». *Mind* 117 (468): 843-65.
- Feinberg, Joel. 1970. *Doing & Deserving: Essays in the Theory of Responsibility*. Princeton: Princeton University Press.
- Fischer, John Martin. 1994. *The Metaphysics of Free Will: An Essay on Control*. Blackwell.
- . 1997. « Responsibility, Control, and Omissions ». *The Journal of Ethics* 1 (1): 45-64. doi:10.1023/A:1009707919608.
- . 2000. « Responsibility, History and Manipulation ». *The Journal of Ethics* 4 (4): 385-91.
- . 2006. *My Way: Essays on Moral Responsibility*. Oxford University Press.
- . 2007. « Compatibilism ». In *Four Views on Free Will*. Wiley-Blackwell.
- . 2010. « The Frankfurt Cases: The Moral of the Stories ». *Philosophical Review* 119 (3): 315-36. doi:10.1215/00318108-2010-002.
- . 2012. « Semicompatibilism and Its Rivals ». *The Journal of Ethics* 16 (2): 117-143. doi:10.1007/s10892-012-9123-9.
- . 2013. « The deterministic horn of the dilemma defence: a reply to Widerker and Goetz ». *Analysis* 73 (3): 489-496. doi:10.1093/analys/ant036.
- . 2016. « How Do Manipulation Arguments Work? » *The Journal of Ethics* 20 (1-3): 47-67. doi:10.1007/s10892-016-9225-x.
- Fischer, John Martin, et Mark Ravizza. 1993. « Responsibility for consequences ». In *Perspectives on Moral Responsibility*, édité par J. Fischer M. Ravizza. Cornell University Press.
- . 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press. <http://ebooks.cambridge.org/ref/id/CBO9780511814594>.
- . , éd. 2000. « Précis of Responsibility and Control: A Theory of Moral Responsibility ». *Philosophy and Phenomenological Research* 61 (2): 441-45. doi:10.2307/2653660.

- Fischer, John Martin, et Neal A. Tognazzini. 2011. « The Physiognomy of Responsibility ». *Philosophy and Phenomenological Research* 82 (2): 381–417.
- Frankfurt, Harry G. 1969. « Alternate Possibilities and Moral Responsibility ». *The Journal of Philosophy* 66 (23): 829–39. doi:10.2307/2023833.
- . 1971. « Freedom of the Will and the Concept of a Person ». *The Journal of Philosophy* 68 (1): 5–20. doi:10.2307/2024717.
- . 2002. *Contours of Agency: Essays on Themes from Harry Frankfurt*. Édité par Sarah Buss et Lee Overton. MIT Press.
- . 2003. « Some thoughts concerning PAP ». In *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, édité par David Widerker et Michael McKenna, 339–345. Ashgate.
- Franklin, Christopher Evan. 2015. « Everyone Thinks That an Ability to Do Otherwise Is Necessary for Free Will and Moral Responsibility ». *Philosophical Studies* 172 (8): 2091–2107. doi:10.1007/s11098-014-0399-4.
- Ginet, Carl. 1966. « Might We Have No Choice? » In *Freedom and Determinism*, édité par Keith Lehrer, 87–104. Random House.
- . 1990. *On Action*. Cambridge University Press.
- . 1996. « In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing ». *Philosophical Perspectives* 10: 403–17. doi:10.2307/2216254.
- . 2006. « Working with Fischer and Ravizza's Account of Moral Responsibility ». *The Journal of Ethics* 10 (3): 229–53. doi:10.1007/s10892-005-5782-0.
- Goetz, Stewart. 2005. « Frankfurt-Style Counterexamples and Begging the Question ». *Midwest Studies in Philosophy* 29 (1): 83–105. doi:10.1111/j.1475-4975.2005.00107.x.
- Haas, Daniel. 2013. « Merit, fit, and basic desert ». *Philosophical Explorations* 16 (2): 226–39. doi:10.1080/13869795.2013.786971.
- Haji, Ishtiyaque. 1992. « A Riddle regarding Omissions ». *Canadian Journal of Philosophy* 22 (4): 485–502.
- . 2009. *Incompatibilism's Allure: Principal Arguments for Incompatibilism*. Critical Issues in Philosophy. Peterborough, Ontario: Broadview Press.
- . 2016. « Luck's Extended Reach ». *The Journal of Ethics* 20 (1–3): 191–218. doi:10.1007/s10892-016-9224-y.
- Haji, Ishtiyaque, et Michael McKenna. 2006. « Defending Frankfurt's Argument in Deterministic Contexts: A Reply to Palmer ». *The Journal of Philosophy* 103 (7): 363–72.
- Hart, H.L.A. 2008. *Punishment and Responsibility*. Édité par John Gardner. Oxford University Press.

<http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199534777.001.0001/acprof-9780199534777>.

Hieronymi, Pamela. 2004. « The Force and Fairness of Blame ». *Philosophical Perspectives* 18: 115-48.

Hume, David. 1739. « 1739. A treatise of human nature ». *London: John Noon*.

———. 1748. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Édité par Nidditch. Oxford: Oxford University Press.

Hunt, David P. 2007. « Black the Libertarian ». *Acta Analytica* 22 (1): 3-15. doi:10.1007/BF02866207.

Hyman, John. 2013. « Voluntariness and Choice ». *The Philosophical Quarterly* 63 (253): 683-708. doi:10.1111/1467-9213.12074.

———. 2015. *Action, Knowledge, and Will*. Oxford: Oxford University Press.

Inwagen, Peter van. 1983. *An Essay on Free Will*. Oxford University Press.

———. 1999. « Moral Responsibility, Determinism, and the Ability to Do Otherwise ». *The Journal of Ethics* 3 (4): 343-51. doi:10.1023/A:1009832417573.

———. 2000. « The Eighth Philosophical Perspectives Lecture: Free Will Remains a Mystery ». *Noûs* 34 (octobre): 1-19. doi:10.1111/0029-4624.34.s14.1.

———. 2008. « How to Think about the Problem of Free Will ». *The Journal of Ethics* 12 (3-4): 327-41. doi:10.1007/s10892-008-9038-7.

Jacobson, Daniel. 2013. « Regret, Agency, and Error ». In *Oxford Studies in Agency and Responsibility. Volume 1 [Electronic Resource]*, édité par David Shoemaker. Oxford Scholarship Online. Oxford: Oxford University Press. <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://dx.doi.org/10.1093/acprof:oso/9780199694853.001.0001>.

Kane, Robert. 1996. *The Significance of Free Will*. Oxford University Press. <http://www.oxfordscholarship.com/view/10.1093/0195126564.001.0001/acprof-9780195126563>.

———. 2016. « Moral Responsibility, Reactive Attitudes and Freedom of Will ». *The Journal of Ethics* 20 (1-3): 229-46. doi:10.1007/s10892-016-9234-9.

Kant, Immanuel. 1781. *Critique de la raison pure*. Vol. 1. Germer-Baillière.

Kelly, Erin. 2008. « 14: Doing without Desert ». In *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's « Freedom and Resentment »*, Ashgate, 273-95. Aldershot: McKenna, Michael; Russell, Paul. <http://web.a.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzl3MDkyM19fQU41?sid=12422570-d0a1-4037-a4cd-0879b6a6deed@sessionmgr4002&vid=0&format=EB&rid=1>.

Korsgaard, Christine M. 1992. « Creating the kingdom of ends: Reciprocity and responsibility in personal relations ». *Philosophical perspectives* 6: 305-32.

Lehrer, Keith. 2016. « Freedom of Preference: A Defense of Compatibilism ». *Journal of Ethics* 20 (1-3): 35-46.

Levy, Neil. 2005. « The Good, the Bad and the Blameworthy ». *Journal of Ethics & Social Philosophy* 1: 1-16.

———. 2011. « Expressing Who We Are: Moral Responsibility and Awareness of Our Reasons for Action ». *Analytic Philosophy* 52 (4): 243-61. doi:10.1111/j.2153-960X.2011.00543.x.

———. 2014. *Consciousness and Moral Responsibility*. Oxford University Press. <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198704638.001.0001/acprof-9780198704638>.

Lewis, David. 1981. « Are We Free to Break the Laws? » *Theoria* 47 (3): 113-21. doi:10.1111/j.1755-2567.1981.tb00473.x.

Locke, John. 1836. *An essay concerning human understanding*. T. Tegg and Son.

Lycan, William G. 1987. *Consciousness*. Bradford Book. Cambridge, Mass ; London: MIT Press.

McKenna, Michael. 2006. « Guest Editor's Introduction ». *The Journal of Ethics* 10 (3): 211-27. doi:10.1007/s10892-005-5778-9.

———. 2008. « Putting the Lie on the Control Condition for Moral Responsibility ». *Philosophical Studies* 139 (1): 29-37. doi:10.1007/s11098-007-9100-5.

———. 2012. « Moral Responsibility, Manipulation Arguments, and History: Assessing the Resilience of Nonhistorical Compatibilism ». *The Journal of Ethics* 16 (2): 145-74. doi:10.1007/s10892-012-9125-7.

———. 2013. « Reasons-Responsiveness, Agents, and Mechanisms ». In *Oxford Studies in Agency and Responsibility. Volume 1 [Electronic Resource]*, édité par David Shoemaker. Oxford Scholarship Online. Oxford: Oxford University Press. <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://dx.doi.org/10.1093/acprof:oso/9780199694853.001.0001>.

———. 2016. « A Modest Historical Theory of Moral Responsibility ». *The Journal of Ethics* 20 (1-3): 83-105. doi:10.1007/s10892-016-9227-8.

McKenna, Michael, et D. Justin Coates. 2015. « Compatibilism ». In *The Stanford Encyclopedia of Philosophy*, édité par Edward N. Zalta, Summer 2015. <http://plato.stanford.edu/archives/sum2015/entries/compatibilism/>.

Mele, Alfred R. 1995. *Autonomous Agents: From Self-Control to Autonomy*. Oxford University Press.

———. 2006a. *Free Will and Luck*. Oxford University Press.

———. 2006b. « Fischer and Ravizza on Moral Responsibility ». *The Journal of Ethics* 10 (3): 283-94. doi:10.1007/s10892-005-5780-2.

———. 2009. « Moral Responsibility and Agents' Histories ». *Philosophical Studies* 142 (2): 161-81. doi:10.1007/s11098-007-9181-1.

———. 2016. « Moral Responsibility: Radical Reversals and Original Designs ». *The Journal of Ethics* 20 (1-3): 69-82. doi:10.1007/s10892-016-9222-0.

Moya, Carlos J. 2011. « On the Very Idea of a Robust Alternative ». *Crítica: Revista Hispanoamericana de Filosofía* 43 (128): 3-26.

Nagel, Thomas. 2012. *Mortal Questions*. Cambridge: Cambridge University Press. <http://ebooks.cambridge.org/ref/id/CBO9781107341050>.

Nelkin, Dana Kay. 2014. « Free Will Skepticism and Obligation Skepticism: Comments on Derk Pereboom's Free Will Agency, and Meaning in Life ». *Science, Religion & Culture* 1 (3): 209-17.

———. 2016. « Accountability and Desert ». *The Journal of Ethics* 20 (1-3): 173-89. doi:10.1007/s10892-016-9230-0.

Nozick, Robert. 1997. « Coercion ». In *Socratic Puzzles*. Harvard University Press.

O. Brink, David, et Dana Kay Nelkin. 2013. « Fairness and the Architecture of Responsibility ». In *Oxford Studies in Agency and Responsibility. Volume 1 [Electronic Resource]*, édité par David Shoemaker. Oxford Scholarship Online. Oxford: Oxford University Press. <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://dx.doi.org/10.1093/acprof:oso/9780199694853.001.0001>.

O'Connor, Timothy. 2016. « Free Will ». In *The Stanford Encyclopedia of Philosophy*, édité par Edward N. Zalta, Summer 2016. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2016/entries/freewill/>.

Palmer, David. 2005. « New Distinctions, Same Troubles: A Reply to Haji and McKenna ». *The Journal of Philosophy* 102 (9): 474-82.

———. 2014. « Deterministic Frankfurt Cases ». *Synthese* 191 (16): 3847-64. doi:10.1007/s11229-014-0500-8.

Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge University Press.

———. 2003. « Source incompatibilism and alternative possibilities ». In *Freedom, Responsibility, and Agency: Essays on the Importance of Alternative Possibilities*, édité par Michael S. McKenna et David Widerker, 184--199. Ashgate.

———. 2007. « Hard incompatibilism ». In *Four Views on Free Will*, édité par John Martin Fischer. Blackwell.

———. 2008. « 11: Revising Reactive Attitudes ». In *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's « Freedom and Resentment »*, Ashgate, 219-34. Aldershot: McKenna, Michael; Russell, Paul. <http://web.a.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzl3MDkyM19fQU41?sid=12422570-d0a1-4037-a4cd-0879b6a6deed@sessionmgr4002&vid=0&format=EB&rid=1>.

———. 2009. « Hard Incompatibilism and Its Rivals ». *Philosophical Studies* 144 (1): 21-33. doi:10.1007/s11098-009-9371-0.

———. 2012. « Frankfurt Examples, Derivative Responsibility, and the Timing Objection ». *Philosophical Issues* 22 (1): 298-315. doi:10.1111/j.1533-6077.2012.00231.x.

———. 2013a. « Free Will Skepticism and Criminal Punishment ». In *The Future of Punishment*, Oxford University Press, 49-78. New York ; Oxford: Thomas Nadelhoffer.

———. 2013b. « Skepticism About Free Will ». In *Exploring the Illusion of Free Will and Moral Responsibility*, édité par Gregg D. Caruso, 19. Lexington Books.

———. 2014. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

Rawls, John. 1980. « Kantian constructivism in moral theory ». *The journal of philosophy* 77 (9): 515-72.

Russell, Paul. 2008. « Strawson's Way of Naturalizing Moral Responsibility ». In *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's « Freedom and Resentment »*, Ashgate, 143-56. Aldershot: McKenna, Michael; Russell, Paul. <http://web.a.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzl3MDkyM19fQU41?sid=12422570-d0a1-4037-a4cd-0879b6a6deed@sessionmgr4002&vid=0&format=EB&rid=1>.

———. 2013. « Responsibility, Naturalism, and "The Morality System" ». In *Oxford Studies in Agency and Responsibility. Volume 1 [Electronic Resource]*, édité par David Shoemaker. Oxford Scholarship Online. Oxford: Oxford University Press. <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://dx.doi.org/10.1093/acprof:oso/9780199694853.001.0001>.

Sartorio, Carolina. 2005. « A New Asymmetry Between Actions And Omissions* ». *Noûs* 39 (3): 460-82. doi:10.1111/j.0029-4624.2005.00509.x.

———. 2015. « Sensitivity to Reasons and Actual Sequence ». In *Oxford Studies in Agency and Responsibility. Volume 3 [Electronic Resource]*, édité par David Shoemaker, First edition. Oxford Scholarship Online. Oxford: Oxford University Press. <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://dx.doi.org/10.1093/acprof:oso/9780198744832.001.0001>.

———. 2016a. *Causation and Free Will*. Oxford: Oxford University Press. <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198746799.001.0001/acprof-9780198746799>.

———. 2016b. « Ignorance, Alternative Possibilities, and the Epistemic Conditions for Responsibility ». In *Perspectives on Ignorance from Moral and Social Philosophy*, édité par Rik Peels, 15-29. Routledge.

Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge, Mass. ; London: Belknap Press of Harvard University Press.

———. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, Mass. ; London: Belknap.

———. 2013. « Giving desert its due ». *Philosophical Explorations* 16 (2): 101-16. doi:10.1080/13869795.2013.787437.

Shoemaker, David, et Neal A. Tognazzini. 2015. *Oxford Studies in Agency and Responsibility. Volume 2, Freedom and Resentment at 50 [Electronic Resource]*. Édité par Oxford University Press. Oxford Scholarship Online. Oxford: Oxford University Press. <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://dx.doi.org/10.1093/acprof:oso/9780198722120.001.0001>.

Shoemaker, David W. 2011. « Moral Responsibility and the Self », février. <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199548019.001.0001/oxfordhb-9780199548019-e-22>.

Sie, Maureen. 2000. « Freedom and Blameworthiness ». In *Moral Responsibility and Ontology*, Kluwer, 113-28. van Den Beld, A. <http://philpapers.org/rec/SIEFAB>.

———. 2005. « Ordinary Wrongdoing and Responsibility Worth Wanting ». *European Journal of Analytic Philosophy* 1 (2): 67-82.

Smart, J. J. C. 1961. « Free Will, Praise, and Blame ». *Mind* 70: 291-306.

———. 1973. « An outline of a system of utilitarian ethics ». In *Utilitarianism: For and Against*, 1-74. Cambridge: Cambridge University Press. <https://www.cambridge.org/core/books/utilitarianism/an-outline-of-a-system-of-utilitarian-ethics/8DE8362FF43188D53C855A70C70223E2>.

Smith, Angela M. 2005. « Responsibility for Attitudes: Activity and Passivity in Mental Life ». *Ethics* 115 (2): 236-71. doi:10.1086/426957.

———. 2007. « On Being Responsible and Holding Responsible ». *The Journal of Ethics* 11 (4): 465-84. doi:10.1007/s10892-005-7989-5.

———. 2008. « Control, Responsibility, and Moral Assessment ». *Philosophical Studies* 138 (3): 367-92. doi:10.1007/s11098-006-9048-x.

Smith, Michael. 2003. « Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion ». In *Weakness of Will and Practical Irrationality*. Oxford: Oxford University Press. <http://www.oxfordscholarship.com/view/10.1093/0199257361.001.0001/acprof-9780199257362-chapter-2>.

Sripada, Chandra. 2015. « Moral Responsibility, Reasons, and the Self ». In *Oxford Studies in Agency and Responsibility. Volume 3 [Electronic Resource]*, édité par David Shoemaker, First edition. Oxford Scholarship Online. Oxford: Oxford University Press. <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://dx.doi.org/10.1093/acprof:oso/9780198744832.001.0001>.

———. 2016. « Self-Expression: A Deep Self Theory of Moral Responsibility ». *Philosophical Studies* 173 (5): 1203-32. doi:10.1007/s11098-015-0527-9.

Strawson, Galen. 1989. « Consciousness, free will, and the unimportance of determinism ». *Inquiry* 32 (1): 3-27. doi:10.1080/00201748908602175.

———. 1994. « The Impossibility of Moral responsibility ». *Philosophical Studies*, n° 75: 5-24.

———. 2010. *Freedom and Belief*. Oxford: Oxford University Press. <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199247493.001.0001/acprof-9780199247493>.

Strawson, P.F. 1974. *Freedom and Resentment and Other Essays*. Routledge.

———. 1985. *Skepticism and Naturalism : Some Varieties*. Woodbridge Lectures ; 1983. London: Methuen.

———. 1998. « Reply to Andrew Black ». In *The Philosophy of P.F. Strawson*, édité par Lewis Edwin Hahn, 242-44. Library of Living Philosophers ; 26. Chicago, Ill: Open Court.

———. 2008. *Freedom and resentment and other essays*. Routledge.

Strawson, P.F., Pranab Kumar Sen, et Roop Rekha Verma. 1995. *The Philosophy of P.F. Strawson*. Distributed by Allied Publishers.

Taylor, Christopher, et Daniel C. Dennett. 2011. « Who's Still Afraid of Determinism? Rethinking Causes and Possibilities ». In *The Oxford Handbook of Free Will*. Oxford University Press. [//www.oxfordhandbooks.com/10.1093/oxfordhb/9780195399691.001.0001/oxfordhb-9780195399691-e-11](http://www.oxfordhandbooks.com/10.1093/oxfordhb/9780195399691.001.0001/oxfordhb-9780195399691-e-11).

Todd, Patrick. forthcoming. « A Unified Account of the Moral Standing to Blame ». *Noûs*.

———. 2011. « A new approach to manipulation arguments ». *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 152 (1): 127-33.

———. 2013. « Defending (a Modified Version of) the Zygote Argument ». *Philosophical Studies* 164 (1): 189-203. doi:10.1007/s11098-011-9848-5.

———. 2016. « Manipulation Arguments and the Freedom to Do Otherwise ». *Philosophy and Phenomenological Research*, n/a-n/a. doi:10.1111/phpr.12298.

Van Inwagen, Peter. 2017. *Thinking about Free Will*. Cambridge: Cambridge University Press.

Vargas, Manuel R. 2005. « The Revisionist's Guide to Responsibility ». *Philosophical Studies* 125 (3): 399-429. doi:10.1007/s11098-005-7783-z.

———. 2006. « On the Importance of History for Responsible Agency ». *Philosophical Studies* 127 (3): 351-82. doi:10.1007/s11098-004-7819-9.

———. 2009. « Revisionism about Free Will: A Statement & Defense ». *Philosophical Studies* 144 (1): 45-62. doi:10.1007/s11098-009-9366-x.

———. 2015. « Précis of Building Better Beings: A Theory of Moral Responsibility ». *Philosophical Studies* 172 (10): 2621-23. doi:10.1007/s11098-015-0476-3.

Vihvelin, Kadri. 2004. « Free Will Demystified: A Dispositional Account ». *Philosophical Topics* 32 (1/2): 427-450.

———. 2013. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oxford University Press. <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199795185.001.0001/acprof-9780199795185>.

———. 2015. « Arguments for Incompatibilism ». In *The Stanford Encyclopedia of Philosophy*, édité par Edward N. Zalta, Fall 2015. <http://plato.stanford.edu/archives/fall2015/entries/incompatibilism-arguments/>.

Wallace, R Jay. 1994. *Responsibility and the moral sentiments*. Harvard University Press.

Wallace, R. Jay. 2002. « Précis of Responsibility and the Moral Sentiments ». *Philosophy and Phenomenological Research* 64 (3): 680-81. doi:10.1111/j.1933-1592.2002.tb00172.x.

Waller, Bruce N. 1990. *Freedom Without Responsibility*. Temple University Press.

———. 2014. *The Stubborn System of Moral Responsibility*. MIT Press. <https://books.google.com/books?hl=fr&lr=&id=QebUBQAAQBAJ&oi=fnd&pg=PR5&dq=the+stubborn+system+of+moral+responsibility&ots=8TYrSkbW42&sig=ulW49KxFeldeFMggX5NJOf6cX0>.

Watson, Gary. 1975. « Free Agency ». *The Journal of Philosophy* 72 (8): 205-20. doi:10.2307/2024703.

———. 1987. « Free Action and Free Will ». *Mind* 96 (382): 145-72.

———. 1996. « Two Faces of Responsibility ». *Philosophical Topics* 24 (2): 227-248. doi:10.5840/philtopics199624222.

———. 1999. « Soft Libertarianism and Hard Compatibilism ». *The Journal of Ethics* 3 (4): 353-68. doi:10.1023/A:1009819618482.

———. 2004. *Agency and Answerability: Selected Essays*. Oxford University Press.

———. 2008. « 6: Responsibility and the Limits of Evil ». In *Free Will and Reactive Attitudes : Perspectives on P.F. Strawson's « Freedom and Resentment »*, Ashgate, 115-41. Aldershot: McKenna, Michael; Russell, Paul. <http://web.a.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzl3MDkyM19fQU41?sid=12422570-d0a1-4037-a4cd-0879b6a6deed@sessionmgr4002&vid=0&format=EB&rid=1>.

———. 2012. « La responsabilité et les limites du mal. Variations sur un thème de Strawson ». *Les ateliers de l'éthique / The Ethics Forum* 7 (1): 146-78.

Widerker, David. 1995. « Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities ». *The Philosophical Review* 104 (2): 247-61. doi:10.2307/2185979.

———. 2000. « Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look ». *Philosophical Perspectives* 14: 181-201.

———. 2006. « Libertarianism and the Philosophical Significance of Frankfurt Scenarios ». *The Journal of Philosophy* 103 (4): 163-87. doi:10.5840/jphil2006103433.

Widerker, David, et Stewart Goetz. 2013. « Fischer against the dilemma defence: the defence prevails ». *Analysis* 73 (2): 283-95. doi:10.1093/analys/ant013.

Williams, Bernard. 1981. *Moral Luck: Philosophical Papers, 1973-1980*. Cambridge: Cambridge University Press.

———. 1993. *Shame and Necessity*. Sather Classical Lectures ; 57. Berkeley ; Oxford: University of California Press.

Wittgenstein, Ludwig. 1974. *On Certainty*. Oxford: Blackwell.

Wolf, Susan. 1993. *Freedom within reason*. Oxford University Press.

Zimmerman, David. 2002. « Reasons-Responsiveness and Ownership-of-Agency: Fischer and Ravizza's Historicist Theory of Responsibility ». *The Journal of Ethics* 6 (3): 199-234. doi:10.1023/A:1019561013541.

Zimmerman, Michael J. 2016. « Moral Responsibility and the Moral Community: Is Moral Responsibility Essentially Interpersonal? » *The Journal of Ethics* 20 (1-3): 247-63. doi:10.1007/s10892-016-9233-x.

Simon-Pierre Chevarie-Cossette

Oxford University

spccossette@gmail.com