

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LE RÔLE DE LA *SYMPATHIE* DANS LA CORRECTION DES JUGEMENTS

MORAUX

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN PHILOSOPHIE

PAR

MARINA DE SETA

MARS 2016

REMERCIEMENTS

Ce mémoire a été l'objet principal de mon attention pendant deux longues années et, comme chaque entreprise que l'on a le désir d'accomplir, il m'a rapproché de certaines personnes tout en m'éloignant de d'autres.

En ce qui concerne les premières, j'ai eu la chance de travailler avec Julie Walsh : une directrice inspirante et une magnifique personne que je peux ici remercier pour sa patience et la motivation qu'elle a réussi à me transmettre. J'ai aussi pu profiter de l'aide importante de mon co-directeur, Luc Faucher, qui a toujours réussi à transformer ce qui a mes yeux était une étape difficile à surmonter, en petite épreuve tout à fait réalisable. Ce faisant, il m'a toujours offert des commentaires formateurs et de suggestions qui poussaient la recherche plus en profondeur. Merci Luc. Pierre Poirer et Dario Perinetti se sont toujours montrés disponibles pour écouter mes doutes et répondre à mes questions. Merci à vous deux pour les discussions qu'on a eues ensemble et pour les conseils que vous m'avez donnés : c'était précieux. Merci aussi à d'autres professeur.e.s qui m'ont fait percevoir l'université comme une place moins dure et distante que lorsque je suis arrivée, entre autres je remercie particulièrement Mauro Rossi, Sara Magrin et Serge Robert.

Un grand merci va aussi aux personnes que je n'aie pas pu voir autant que je l'aurais souhaité durant ces deux années de rédaction, mais qui m'ont malgré cela accompagné lors de mon parcours. Je tiens à remercier Clélia Sève, Renaud Santerre Crête, Zoé Schoen et Alexandre Duval : vos commentaires, vos questions, vos critiques et vos doutes concernant ce projet m'ont donné de la motivation et des raisons pour vouloir l'améliorer. Je remercie aussi toutes les filles de l'équipe de *Filosophie* qui ont été une raison constante d'implication dans le département. Votre énergie et votre envie de faire est inspirante.

Des remerciements et des excuses vont à ceux qui m'ont vu disparaître de tous les réseaux sociaux sans avoir d'autres moyens de me contacter. Giulia Pasquini,

Ilaria Aguzzi, Flavia Cordone, Sara Giuliano, Giulio Cancellara, Edoardo Nagali et Luca Esposito: merci de ne pas m'avoir laissée seule malgré mes horaires chargés et le décalage. Votre amitié est et sera toujours précieuse pour moi, peu importe où nous serons sur la planète (ou en dehors, on sait jamais!). Sans vous, je ne serais pas la personne que je suis, et j'aurais très probablement perdu la tête avant de terminer ce mémoire.

Enfin, des simples remerciements ne sont pas suffisants pour toutes les gens que je considère ma famille. Tout d'abord, mes grands-parents, Anna Ceccarelli et Fernando Scotti : vous avez toujours inspiré le côté fort et déterminé qui souvent me motive à continuer et finir ce que j'ai commencé. Pour vous montrer, et prouver à moi-même, que j'en suis capable, j'ai osé sortir de ma zone de confort, et je ne le regretterai jamais. Grâce à vos enseignements j'ai pu me distinguer et grâce à vos questionnements et à votre esprit critique, je suis devenue curieuse. Un grand merci va à vous deux. Grand-papa, je n'ai jamais oublié qu'il vaut mieux être un âne stupide qu'un sage mort. C'est devenu ma devise.

Ensuite, je remercie Rita, Anna, Margherita et Mario : vous avez été ma famille et mon refuge. Votre grand cœur a permis que je grandisse gaie et sans soucis. Ce mémoire est aussi à vous.

Je remercie ma mère et mon père : vous êtes la raison de tout ce que j'ai pu obtenir de cette vie. Je crois que les mots ne peuvent pas exprimer ma gratitude. J'ai eu la chance de naître dans une maison stimulante, riche en discussions et entourée de personnes différentes. Vous m'avez permis de voyager et d'expérimenter et vous êtes toujours là lors des moments difficiles. Ma chance est énorme, et c'est souvent le fruit de vos efforts. Pour cela, je ne pourrais jamais assez vous remercier.

Alex, ce mémoire ne serait pas fini sans toi. Merci d'avoir toléré que j'aie envie de discuter de mon sujet de mémoire à des horaires impossibles ou dans des circonstances peu favorables. Merci d'avoir écouté, cela nonobstant. Merci d'avoir lu, modifié, corrigé et relu toutes les sections de ce mémoire. Merci de ta

patience et de ton support, même lorsque nous n'étions pas dans le même coin du monde. Ta présence dans ma vie me motive chaque jour. Je te promets que pour au moins quelque mois je serai une personne adorable et agréable à nouveau. Cette aventure est presque conclue mais nous en aurons tellement d'autres à vivre, donc tiens-toi prêt.

L'écriture de ce mémoire a été plus longue qu'espéré et elle a été mise à l'épreuve par plusieurs changements et événements. Le dernier remerciement va donc à moi-même. Nonobstant tout ça, t'as réussi. Bravo !

TABLE DES MATIÈRES

REMERCIEMENTS	ii
TABLE DES MATIÈRES	v
LISTE DES TABLEAUX.....	viii
LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES	ix
RÉSUMÉ	xi
INTRODUCTION	1
0.1 Deux perspectives sur les jugements moraux	4
0.2 La structure de ce mémoire	7
CHAPITRE I.....	12
LA RÉVISION DES JUGEMENTS MORAUX : UNE PERSPECTIVE PSYCHOLOGIQUE	12
1.1 Un esprit biface.....	12
1.1.1 Le modèle de Kahneman	13
1.1.2 Comment écarter un problème des théories des systèmes duaux	15
1.1.3 La collaboration entre les deux systèmes	18
1.1.4 Flexibilité et apprentissage des deux systèmes.....	20
1.2 Une solution possible	21
1.2.1 Favoriser les raisonnements conséquentialistes.....	22
1.2.2 <i>Trolleyologie</i>	23
1.2.3 Le modèle de Greene	30
1.2.4 Les limites de la méthode proposée par Greene	32
1.3 Le conducteur et l'éléphant	35
1.3.1 Le modèle socio-intuitionniste de Haidt	35

1.3.2	Les PT2 : non seulement des réflexions <i>post-hoc</i>	40
1.3.3	Réviser un jugement	45
1.4	Conclusion	52
CHAPITRE II		54
UNE SOLUTION HUMIENNE		54
2.1	L'objet du jugement moral.....	56
2.1.1	L'association d'idées	58
2.1.2	Le mécanisme de la sympathie	65
2.1.3	Le contenu des jugements moraux.....	70
2.2	La sympathie limitée et les erreurs de jugement.....	76
2.2.1	Percevoir les passions d'autrui	76
2.2.2	Le caractère.....	80
2.2.3	Les « erreurs » de la sympathie limitée	83
2.3	La correction des sentiments inappropriés.....	85
2.3.1	Pourquoi atteindre le PVC?	87
2.3.2	Le PVC	90
2.3.3	Le PVC et la sympathie étendue.....	92
2.3.4	Le contenu du jugement corrigé : les sentiments moraux	94
2.3.5	L'idéal moral humien : le spectateur judicieux	96
CONCLUSION.....		104
BIBLIOGRAPHIE		109

LISTE DES FIGURES

Figure		Page
1.1.3	Les trois types d'interaction entre S1 et S2 dans le modèle de Kahneman.....	18
1.2.3	L'activité des PT1 et des PT2 dans la formulation des jugements moraux en relation au temps.....	31
1.3.1	Traduction du schéma du modèle socio-intuitionniste (Jonathan Haidt 2012, p. 55).....	38
1.3.3	PT1 et PT2 dans le modèle socio-intuitionniste de Haidt.....	44
1.3.3.1	Les processus de correction des jugements moraux dans le modèle socio-intuitionniste.....	46
2.1.1	Les types de perceptions dans la théorie psychologique de David Hume.....	59
2.1.2	L'influence des relations des idées et de leur association sur la sympathie.....	69
2.1.3	Les vices, les vertus et les passions indirectes	71
2.3.3	La sympathie étendue implique une variation des relations entre les idées.....	93

LISTE DES TABLEAUX

Tableau		Page
1.1.1	Traduction partielle du Tableau à p. 51 de Kahneman & Frederick 2002.....	13
1.3.1	Les caractéristiques des processus duaux impliqués dans l'élaboration des jugements moraux (Jonathan Haidt)....	36

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

David Hume, *Traité de la Nature Humaine* :

- (T 1.2.3.4) Renvoie au Livre I, Partie II, Section 3, Paragraphe 4 de l'édition anglaise de David Fate Norton et Mary J. Norton, Oxford, Clarendon Press, 2007. Les citations sont tirées de la traduction de P. Baranger et P. Saltel, éditée par Flammarion (1991, 1993, 1995).

David Hume, *Enquête sur l'Entendement Humain* :

- (EEH 4.1.2) Renvoie à la Section 4, Partie 1, Paragraphe 2 de l'édition anglaise de Tom L. Beauchamp, (The Clarendon Edition of the Works of David Hume), Oxford, Oxford University Press, 1998. (Lorsque la référence contient seulement deux chiffres, le premier renvoie à la *Section* et le deuxième au *Paragraphe*). Les citations sont tirées de la traduction de A. Leroy et M. Beyssade éditée par Flammarion (2006).

David Hume, *Enquête sur les Principes de la Morale* :

- (EPM 3.2.6) Renvoie à la Section 3, Partie 2, Paragraphe 6 de l'édition anglaise de Tom L. Beauchamp, Oxford, Oxford University Press, 1999. (Lorsque la référence contient seulement deux chiffres, le premier renvoie à la *Section* et le deuxième au *Paragraphe*). Les citations sont tirées de la traduction de P. Baranger et P. Saltel, éditée par Flammarion (1991).

David Hume, *Essai sur la Norme du Goût* :

- (ENG 22) Renvoie au Paragraphe 22 de l'édition anglaise des *Essays Moral, Political, and Literary* d'Eugene F. Miller, Indianapolis: Liberty Classics, 1985. Les citations sont tirées de la traduction de R. Bouveresse éditée par Flammarion (2000).

David Hume, *Des Caractères Nationaux* :

(NC 9) Renvoie au Paragraphe 9 de l'édition anglaise des *Essays Moral, Political, and Literary* d'Eugene F. Miller, Indianapolis: Liberty Classics, 1985.

RÉSUMÉ

Ce mémoire porte sur les types de méthodes qui peuvent être employées pour corriger nos jugements moraux biaisés. Le point de départ de cette recherche est l'hypothèse selon laquelle notre capacité à formuler des jugements moraux joue un rôle crucial dans la formation et le maintien du tissu social. Ce travail cherche à identifier quels processus cognitifs interviennent lorsque nous formulons des jugements moraux et comment nous pourrions diminuer l'influence des biais cognitifs sur eux. Dans ce but, nous présentons deux des modèles de jugement moral les plus influents dans la littérature contemporaine, celui de Joshua Greene et celui de Jonathan Haidt. Nous concluons de cette présentation que les modèles ont des problèmes (chap. 1). Conséquemment, nous proposons une méthode de monitoring des jugements moraux inspirée de la théorie morale de David Hume (chap. 2). En suivant ses pas, nous suggérons que notre capacité à changer de point de vue apporte une solution au problème de l'influence des biais cognitifs sur les jugements moraux.

Mots clés : David Hume; Joshua Greene; Jonathan Haidt; naturalisme; processus-duaux; sympathie; empathie; jugements moraux; prise de rôle; biais affectif.

INTRODUCTION

Les réflexions qui seront présentées dans ce mémoire portent sur les jugements moraux. La littérature contemporaine à ce sujet est très vaste et de nombreuses questions à ce propos demeurent controversées, notamment : qu'est-ce un jugement moral ? Les jugements moraux ont-ils un référent ? Pourquoi est-ce qu'on juge moralement ? Or, la suite de ce texte risque de manquer de clarté si on laisse sans réponse la première de ces questions. En laissant de côté toutes les discussions en méta-éthique qui visent à déterminer plus précisément ce qui est un jugement moral, il suffira de dire qu'un jugement moral est ce qui nous mène à évaluer un individu, son comportement, ses attitudes ou même ses émotions comme *bonnes* ou *mauvaises*. Cela dit, pourquoi devrions-nous nous intéresser aux questions concernant les jugements moraux ? L'anthropologue et psychologue évolutionniste Robin Dunbar et le psychologue Joshua Greene proposent une réponse intéressante à cette question.

Dunbar affirme que « sans le *commérage* [*gossip*], il n'y aurait pas de société » et que c'est ce « qui rend possible la société des êtres humains telle que nous [la] connaissons » (Dunbar 2001, p. 100). Le terme « commérage » est ici dénué de la connotation négative que nous lui donnons communément. Pour Dunbar, ce terme indique plutôt l'échange d'informations qui a lieu lorsqu'on interagit les uns avec les autres pour « (a) suivre les changements et les activités des autres membres de notre réseau (b) pour faire de la publicité pour son propre bénéfice en tant qu'amis, alliés ou partenaires (et peut-être à propos des inconvénients des rivaux potentiels) (c) pour rechercher des conseils concernant des problèmes personnels et, finalement (d) pour surveiller les profiteurs [*free riders*] », à savoir ceux qui tirent avantage de la vie en société sans en payer les coûts (Dunbar 2004, p. 105, énumération ajoutée).

Lorsqu'on accepte de définir le commérage de cette façon, nous constatons que notre capacité à formuler des évaluations morales y figure de façon centrale. C'est

notamment le cas du point (b). En effet, lorsque je dis à quelqu'un que je sais garder un secret, cela implique une évaluation positive de ce trait de mon caractère. Il en va de même du point (c). En conversant avec les autres à propos de nos problèmes personnels, nous leur proposons souvent nos solutions et ils nous offrent les leurs en réponse. Cela peut impliquer que les participants à ce genre de conversation jugent ce qui serait une bonne ou une mauvaise réaction à la situation dont il est question. Quant au point (d), il me semble que le commérage ayant pour but de tenir sous contrôle les « profiteurs » repose sur un jugement de valeur à propos du comportement de ces individus. En d'autres mots, nous les tenons sous contrôle puisque nous les considérons nuisibles.

À la lumière de ce qui précède, nous avons de bonnes raisons de croire que les jugements moraux constituent une partie importante des contenus de nos commérages. Étant donné que pour Dunbar la société est possible grâce au commérage, il en suit que les jugements moraux joueraient un rôle important dans la formation de notre société. Même si j'ai quelques réserves concernant la position de Dunbar qui accorde *tout* le crédit au commérage dans la constitution de la société humaine (Dunbar 2004, p. 100), je crois cependant que nous pouvons affirmer que le commérage, et donc aussi les jugements moraux, jouent un rôle important dans la formation et dans la préservation de notre société.¹ Je vais consacrer quelques paragraphes à l'explication de cette affirmation.

¹ En effet, les résultats des recherches de Dunbar, que j'illustrerai dans les paragraphes suivants, ne lui permettent pas encore de soutenir que la formation de notre société dépend *principalement* de notre capacité à faire du commérage. Dunbar formule l'hypothèse selon laquelle notre capacité à faire du commérage a substitué un important outil de socialisation chez les singes, à savoir le toilettage mutuel. Cela peut lui permettre d'affirmer que le commérage joue un rôle important dans la formation de la société, sans exclure que d'autres comportements sociaux puissent avoir joué un rôle également ou même plus important dans la formation de la société humaine (par exemple, la communication visant à transmettre des pratiques utiles pour la survie).

Le point de départ de l'argument de Dunbar est le suivant : la société est utile pour la survie seulement si nous nous assurons que chacun de ses membres (ou du moins une bonne partie) sacrifie une partie de son propre profit pour le bien général lorsque nécessaire. Dunbar affirme que, chez les singes, l'obligation de faire passer le bien du groupe avant son propre bien naît des liens affectifs qui sont constitués par le toilettage mutuel, auquel certains singes dévouent environ 20% de leur temps d'éveil (Dunbar 2004, p. 101).²

Selon les recherches conduites par Dunbar, nous dévouons le même pourcentage de notre temps d'éveil non pas au toilettage mutuel, mais à l'interaction sociale (Dunbar 2004, p. 102). Toujours selon lui, dans environ les deux tiers du temps que nous dévouons à l'interaction sociale nous faisons du commérage. Considérant ces résultats, Dunbar formule l'hypothèse selon laquelle le commérage s'est substitué au toilettage mutuel à un certain point de la phylogenèse de l'espèce humaine.³ Ainsi, si le commérage nous permet, entre autres, de créer des liens sociaux, et si, comme je le soutiens, les jugements moraux en constituent une bonne partie, alors les jugements moraux seraient un important outil de socialisation et de création des liens sociaux.

² Cette hypothèse est confirmée par le fait que dans le cas des groupes de primates qui sont très stables (p. ex. ceux des grands singes), plus le groupe est grand, plus le temps dédié au toilettage mutuel est long. Par contre, dans les groupes de primates dont les liens sociaux sont plus instables (p. ex. chez les babouins) le temps dévolu à cette activité est inférieur à 20 % (Dunbar 2004, p. 102).

³ Plus précisément, il suppose que depuis que les hominides ont commencé à constituer des groupes de plus en plus grands afin de mieux se défendre des prédateurs, le temps alloué à créer des liens affectifs par le toilettage mutuel avec chaque membre du groupe s'est rétréci. Le langage est ainsi devenu un mécanisme alternatif pour constituer ces liens affectifs assurant le fonctionnement de la société et établissant la confiance entre ses membres (Dunbar 2004, p. 102). Bien que cette hypothèse me semble acceptable pour le moment, je n'adhère pas aux explications téléologiques de Dunbar pour qui le langage a évolué afin que l'on puisse faire du commérage. Cela ne nous empêche pourtant pas de considérer le langage et le commérage comme étant des instruments importants dans l'apparition des sociétés humaines.

Joshua Greene (2013) formule une hypothèse semblable, en soutenant que la capacité de formuler des jugements moraux nous oriente afin de savoir à qui nous pouvons faire confiance et avec qui nous pouvons collaborer. Selon Greene, c'est à cause de cela que nous sommes devenus biologiquement prédisposés à préférer ceux qu'on considère comme appartenant à nos mêmes « cercles » (social, linguistique, culturel, religieux, sportif, etc.) (Greene, 2013, p.23-24). Toutefois, ce faisant, nous créons un contraste *entre* les groupes, que Greene résume avec la formule « *Nous* versus *Eux* » (*Us* versus *Them*). De nos jours, dans des sociétés composées par plusieurs petits groupes, cette opposition est très souvent cause d'antagonismes, violents ou non.

Les hypothèses de Dunbar et de Greene constituent un point de départ intéressant pour aborder le sujet de ce mémoire. En effet, si nous acceptons que les jugements moraux jouent un rôle de premier plan dans la formation du tissu social, et qu'ils sont à l'origine des préférences subjectives générant des contrastes entre groupes, il devient crucial de bien comprendre (1) comment nous formulons des jugements moraux et (2) comment faire en sorte que ceux-ci ne soient pas l'expression de préférences subjectives.

0.1 Deux perspectives sur les jugements moraux

Au cours des dernières années, plusieurs chercheurs (Haidt 2001 ; Greene 2004 ; Huebner, Dwyer et Hauser 2008) ont soutenu que nos jugements moraux sont déterminés (souvent, voire toujours) par nos intuitions.⁴ D'autres auteurs (Gigerenzer 2008 ; Kahneman & Tversky 1974) soutiennent que les intuitions

⁴ Par « intuition » ces chercheurs désignent une opération mentale inconsciente, rapide et automatique.

jouent un rôle très important dans l'économie de l'esprit, puisqu'elles permettent des réponses aux stimuli qui sont efficaces, rapides et sans effort. Elles nous permettent de réagir à une situation même lorsque nous n'avons pas (toutes) les informations nécessaires pour le faire délibérément. Les intuitions sont supposées être le produit de certaines règles générales, appelées *heuristiques*. L'utilisation de ces dernières peut être très utile mais peut aussi nous induire à l'erreur, et dans ce cas, le résultat de l'heuristique est une intuition biaisée (Gigerenzer 2007, p. 18 ; Kahneman, Slovic et Tversky 1982, p. 4 ; Susteiu 2005, p. 532).⁵

Or, une vaste littérature (Greene 2013 ; Wilson & Brekke 1994) montre que les jugements qui découlent d'intuitions biaisées peuvent être corrigés dans certaines circonstances (lorsque les individus ont du temps, sont motivés, sont conscients de leurs biais, etc.). Toutefois, la correction du jugement demande un effort volontaire plus ou moins important et elle exige donc plus de ressources cognitives (attention, mémoire, etc.) que les intuitions. Rendre intuitifs les processus de révision des jugements moraux serait ainsi profitable pour l'économie et l'efficacité de l'esprit.

Cela dit, il semble maintenant nécessaire d'expliquer selon quel critère nous pouvons affirmer qu'un jugement *moral* est biaisé ou non. Il faut reconnaître qu'il y a des nombreux types de biais différents qui peuvent influencer nos jugements moraux. Conséquemment, le critère que nous pourrions proposer ici pourrait valoir pour les jugements qui découlent de certains biais mais pas pour ceux qui sont influencés par d'autres. En raison de cette variabilité, je juge plus pertinent de prendre en considération principalement le biais affectif [*affective bias*] dans ce mémoire. Ce choix est justifié par le fait que, dans les dernières années, les recherches sur les jugements moraux semblent indiquer que nos intuitions morales

⁵ Le terme biais indique une déviation systématique du jugement, menant un sujet à répondre de façon α à un stimulus α .

ont un contenu affectif (Prinz 2006 ; Maibom 2010 ; Eskine, Kacirik, et Prinz 2011 ; Strohminger, Lewis et Meyer 2011 ; Seidel et Prinz 2013). Suite à ce type de résultats, Slovic soutient que les affects peuvent avoir la même fonction que d'autres heuristiques pour notre esprit. Et il ajoute que :

[...] comme d'autres heuristiques qui fournissent des réponses efficaces et ayant généralement un avantage adaptatif, mais qui nous induisent parfois en erreur, la dépendance [*reliance*] à l'égard de nos affects peut aussi nous tromper (Slovic et al. 2007, p. 1347).

Je crois que ce qui rend une telle heuristique un *biais*, c'est la tournure subjective du jugement qui en suit. Au contraire, les jugements qu'on considère comme ayant été corrigés ont une marque d'objectivité. Je considérerai donc dans ce mémoire que le biais est une déviation systématique du jugement qui nous empêche d'évaluer objectivement la situation dont il est question.

Dans cette perspective, je répondrai principalement, dans ce mémoire, à la question suivante : « Est-il possible d'apprendre à corriger automatiquement nos jugements moraux afin d'éviter qu'ils soient influencés par le biais affectif ? ». Pour ce faire, il faudra répondre préalablement à une autre question, à savoir : « Quelles sont les opérations mentales responsables de la révision de nos jugements moraux ? ». À ce propos, plusieurs publications traitent de l'influence que le changement de perspective peut avoir sur les jugements moraux (Chap. 1986 ; Haidt 2001 ; Eisenberg et al. 2001). Cependant, les auteurs de ces articles n'explorent pas la possibilité d'automatiser cette opération de l'esprit dans le but de corriger plus facilement nos jugements moraux. Dans le sillage de David Hume, je vais soutenir qu'entraîner nos capacités à changer notre point de vue sur la situation pourrait rendre la révision d'un jugement moral plus rapide et plus aisée, en nous permettant même de prévenir les erreurs de jugement.

Hume est généralement connu pour être un *sentimentaliste*, à savoir, qu'il soutiendrait que nos jugements moraux reposent sur un sentiment. Je crois pourtant que la théorie morale de Hume est beaucoup moins axée sur les sentiments que ce que l'on croit. En effet, Hume dit clairement que le sentiment qui constitue le contenu du jugement moral est le fruit de raisonnements (*EPM* 1.9) sur lesquels nous pouvons bien souvent nous tromper. Plus précisément, Hume croit que nos jugements moraux sont biaisés lorsqu'on juge quelqu'un à partir de notre point de vue particulier et subjectif (Debes 2007b, p. 39). Cette perspective limitée sur la personne qu'on est en train de juger nous mène à mésinterpréter ses intentions. Cette erreur de raisonnement nous porte éventuellement à ressentir un sentiment qui sera inapproprié envers cette personne. Afin de corriger les jugements erronés ou même d'éviter de telles erreurs, Hume soutient qu'il faut observer la situation *à partir d'un point de vue commun*.

Selon l'interprétation de la théorie humienne présentée dans ce mémoire, quitter notre point de vue personnel afin d'occuper le point de vue commun nous permettrait de remettre en discussion les croyances et les préjugés qui nous portent à mésinterpréter les intentions des autres. C'est d'ailleurs ce que dit Hume en affirmant que lorsqu'on juge moralement quelqu'un à partir du point de vue commun, nous finissons par ressentir *un sentiment approprié* envers cette personne. Qui plus est, Hume soutient que, grâce à l'habitude, nous pouvons aboutir à adopter le point de vue commun de façon à prévenir les « erreurs » de jugement.

0.2 La structure de ce mémoire

Avant d'entrer dans le vif du sujet, j'aimerais exposer plus en détail les divers moments de la discussion qui aura lieu dans ce mémoire.

Dans le premier chapitre, j'exposerai le plus précisément possible les processus qui entrent en jeu lorsqu'on formule un jugement moral du point de vue de la psychologie morale. Je centrerai la discussion sur les opérations cognitives qui sont responsables de la correction des jugements moraux. À ce propos, je présenterai les deux modèles des jugements moraux les plus influents dans la littérature à ce sujet, à savoir celui conçu par Greene et celui conçu par Jonathan Haidt.

Dans un premier moment (section 1), je décrirai les théories des systèmes duaux puisque les modèles de Greene et de Haidt appartiennent tous deux à cette même grande famille. Selon les défenseurs de ces théories, l'esprit fonctionne grâce à deux systèmes : l'un qui regroupe tous les processus opérant de façon rapide, automatique et sans effort ; l'autre qui réunit les processus lents, qui demandent un effort cognitif plus important et qui sont délibérés. Étant donné la variété dans la formulation des théories des systèmes duaux, je me limiterai à exposer celle conçue par le psychologue Daniel Kahneman (2003) puisque les modèles des jugements moraux de Greene et de Haidt en sont explicitement inspirés.

Dans un deuxième moment (section 2), j'exposerai en détail le modèle de jugement moral proposé par Greene, les expériences sur lesquelles il s'appuie ainsi que la méthode qu'il propose afin de réviser nos jugements moraux. En premier lieu, nous verrons que, suite à ses expériences, Greene considère que nos jugements moraux intuitifs sont déterminés par notre réaction émotionnelle face à la situation jugée. La suggestion de Greene pour diminuer l'influence de ces intuitions sur nos jugements moraux consiste à douter de ceux-ci et à réfléchir davantage. Je montrerai que cette proposition présente trois problèmes. Ces problèmes constituent la raison principale pour laquelle j'estime que la théorie de

Haidt présente des avantages par rapport à celle de Greene. Je soutiendrai qu'il est préférable d'écarter le modèle de Greene pour lui préférer celui de Haidt, modèle que je présenterai donc en dernier.

Dans un troisième moment (section 3), je présenterai le modèle de Haidt en soulignant que, pour lui, nos intuitions morales ont également un contenu affectif et, conséquemment, que les jugements moraux qui en dérivent peuvent être considérés comme étant le produit d'une heuristique de l'affect. Cela dit, je porterai mon attention sur les deux processus cognitifs qui, selon Haidt, peuvent permettre la révision de nos jugements moraux, soit le raisonnement logique et le changement de perspective. Je montrerai que Haidt ne prend pas en considération l'option de rendre automatiques nos capacités de révision des jugements moraux dans son modèle. Ainsi, bien que la théorie de Haidt puisse nous aider à comprendre quels processus cognitifs sont responsables de la correction des jugements moraux, elle ne nous permet pas de répondre à notre interrogation concernant la possibilité de faciliter et de rendre automatique la mise en place de mécanismes de correction des jugements moraux biaisés. Cela ouvrira la porte au deuxième chapitre de ce mémoire où je montrerai comment pour David Hume nous pouvons faciliter la révision de nos jugements moraux en adoptant souvent un changement de point de vue.

Dans le deuxième chapitre de ce mémoire, j'illustrerai comment, selon Hume, nous formulons des jugements moraux et comment nous pouvons les corriger lorsqu'ils sont biaisés. Tel que mentionné auparavant, pour Hume, un jugement moral découle directement d'un sentiment qui découle à son tour de certains raisonnements.

Dans un premier moment (section 1), j'expliquerai quelles sont les trois étapes de la formulation des évaluations morales pour Hume. Premièrement, sa théorie prévoit qu'il faut tout d'abord comprendre quelle intention se cache derrière les actions afin de formuler un jugement moral (de fait, pour Hume les jugements moraux portent sur les intentions et non pas sur l'action en elle-même). Ensuite, j'expliquerai le fonctionnement du mécanisme *sympathique*, puisque de là découlent les sentiments qui constituent le contenu de nos évaluations morales. Enfin, j'argumenterai à l'effet que ces sentiments sont des types particuliers de plaisir et de douleurs selon Hume.

Dans un deuxième moment (section 2), je montrerai quelles sont les erreurs de jugement que nous pouvons commettre dans le cadre de la théorie morale humienne. Entre autres, je montrerai que nous pouvons nous tromper en lisant les intentions des autres, mais aussi en jugeant quelqu'un pour autre chose que son caractère. J'expliquerai enfin comment ces erreurs de jugement découlent de la *sympathie limitée* et comment elle est biaisée en elle-même.

Dans un troisième moment de ce second chapitre (section 3), je montrerai que pour Hume nous pouvons corriger nos jugements moraux en ayant recours à un autre type de sympathie qu'il appelle « sympathie *étendue* ». Cette sympathie repose sur les mêmes principes que la *sympathie limitée*, mais elle demande pour sa part l'aide de l'imagination. En effet, la sympathie est étendue lorsque le sujet imagine d'occuper un point de vue commun lorsqu'il juge moralement. De cette forme de sympathie suivent, selon Hume, les *sentiments appropriés* qui nous permettent de formuler des jugements non biaisés.

Il est maintenant évident que ce mémoire touche non seulement à la philosophie morale mais aussi à la psychologie et aux sciences cognitives. Pour cette raison, il est nécessaire de préciser qu'il y a une grande différence entre les objectifs que Hume voulait atteindre et ceux souhaités par Greene ou par Haidt.

En effet, le but ultime que Hume se proposait d'accomplir avec le *Traité de la Nature Humaine* (dorénavant T ou *Traité*, 1739-40), et aussi avec l'*Enquête sur l'Entendement Humain* (dorénavant EEH, 1748), l'*Enquête sur les Principes de la Morale* (dorénavant EPM, 1751) et la *Dissertation sur les passions* (dorénavant D, 1757) qui sont des approfondissements des Livres I, III et II du *Traité* respectivement, était celui de produire une science de la nature humaine. Dans ces textes, Hume introduit les connaissances concernant la nature humaine qui, à son avis, pourront jouer le rôle de fondements d'une nouvelle organisation des sciences.⁶

Par contre, le but de Greene ou de Haidt est beaucoup plus modeste. Greene vise à : 1) décrire le plus fidèlement possible les processus qui entrent en jeu lorsqu'on formule des jugements moraux et 2) proposer une méthode pour formuler de bons jugements. Haidt se borne à accomplir la première de ces tâches. Malgré les différences quant à la portée de leurs recherches, Hume, Haidt et Greene partagent la même méthode. Tous les trois souhaitent expliquer le fonctionnement de l'esprit humain en partant de l'observation pour en dériver les principes qui le gouvernent et en cela, leur travail en est un de psychologie.⁷

⁶ « Il est évident que toutes les sciences sont plus ou moins reliées à la nature humaine et que, si loin que certaines d'entre elles puissent paraître s'en écarter, elles y reviennent toujours par une voie ou une autre [...] en prétendant expliquer les principes de la nature humaine, nous proposons en fait un système complet des sciences, bâti sur un fondement presque entièrement nouveau, le seul sur lequel elles puissent s'établir avec quelque sécurité » (T Introduction, 4-6).

⁷ Dans ce mémoire, j'aurais recours à plusieurs textes de Hume, à savoir le *Traité de la Nature Humaine*, l'*Enquête sur les Principes de la Morale*, l'*Enquête sur l'Entendement Humain*, et enfin l'*Essai sur la Norme du Goût*, afin d'expliquer de façon satisfaisante sa théorie morale. Pour ce faire, je suppose qu'il y ait une certaine continuité entre les ouvrages que je viens de mentionner, et il y a plusieurs raisons de croire que pour Hume ces textes présentaient une théorie cohérente. Il y a plusieurs preuves textuelles qui justifient d'aborder l'ouvrage de Hume d'une telle manière (Debes 2007a, p. 321-325). En conclusion, dans le but de répondre à la question centrale de ce

CHAPITRE I

LA RÉVISION DES JUGEMENTS MORAUX : UNE PERSPECTIVE PSYCHOLOGIQUE

1.1 Un esprit bifide

Lors des trente dernières années, les chercheurs œuvrant dans divers champs d'études (surtout en psychologie sociale) ont élaboré les *théories des processus duaux* afin d'expliquer le fonctionnement de l'esprit dans ses multiples activités : la mémoire, la prise de décision, les jugements de perception, les jugements moraux et bien d'autres encore (Evans 2008 et 2012 ; Frankish & Evans 2009). Selon les tenants des théories des processus duaux, notre esprit réalise chacune de ces tâches grâce à deux types de processus : les processus de type 1 (dorénavant PT1) et les processus de type 2 (dorénavant PT2).

Certains de ces chercheurs, tels que Daniel Kahneman (2003) et Keith Stanovich (2004), affirment au surplus que l'esprit est divisé en deux larges *systèmes* d'opérations cognitives qui sont des « collections de processus qui se distinguent par la vitesse, la contrôlabilité et les contenus sur lesquels ils opèrent » (Kahneman & Frederick 2002, p. 51). Ainsi, ils soutiennent que la collection des PT1 constitue le Système 1 (dorénavant S1) et que l'ensemble des PT2 constitue le Système 2 (dorénavant S2). Pour cette raison, ces théories sont connues sous le nom de *théories des systèmes duaux*. Les tenants de ces théories soutiennent généralement que, dans de nombreux cas, les réponses de l'esprit aux stimuli sont le produit soit des processus de type 1, soit ceux de type 2, soit de l'activité combinée ou concurrentielle des deux types de processus (qu'ils se trouvent ou

mémoire, je tiendrai pour acquis qu'il y a une certaine continuité au sein des ouvrages de Hume auxquels je ferai référence, tout en sachant qu'ils pourront trouver mention des problèmes interprétatifs plus importants à ce propos tout au long du texte.

non dans deux systèmes différents). Je vais maintenant décrire la théorie des systèmes duaux proposée par le psychologue Daniel Kahneman (2003) qui a inspiré Greene et Haidt, avant de passer à la présentation des théories de ces deux psychologues au sujet des jugements moraux et de leur correction.

1.1.1 Le modèle de Kahneman

Comme nous l'avons mentionné, selon Kahneman les deux processus appartenant aux systèmes cognitifs S1 et S2 se caractérisent par la vitesse et la contrôlabilité de leurs opérations, ainsi que par la différence des contenus qu'ils « manipulent ». Cet auteur énumère d'autres traits typiques de ces deux systèmes cognitifs. Je vais présenter ceux que j'utiliserai plus tard dans mon évaluation des théories des systèmes duaux des jugements moraux.

	Système 1 (Intuitif)	Système 2 (Réflexif)
Caractéristiques des processus	Automatique Sans effort Rapide Inconscient	Contrôlé Avec effort Lent Conscient Application de règles
Contenu des processus	Affectif/ Neutre	Neutre

Tab. 1.1.1 Traduction partielle du tableau à p. 51 de Kahneman & Frederick 2002

Le tableau 1.1.1 montre quelles sont les propriétés des deux types de processus constituant les systèmes duaux.

Tout d'abord, Kahneman caractérise le S1 comme étant intuitif et le S2 comme étant réflexif (tableau 1.1.1). Cela explique notamment pourquoi pour cet auteur, nos jugements peuvent être intuitifs ou réflexifs (Kahneman & Frederik 2002, p. 51) : dans le premier cas, nos jugements sont déterminés par des intuitions; dans le deuxième cas, ils découlent de nos raisonnements.

Selon Kahneman, les PT1 composant le S1 réagissent de façon automatique, sans effort et rapidement aux stimuli. De plus, il affirme que ces processus sont opaques, autrement dit, que nous ne pouvons pas avoir accès par introspection aux opérations qui ont produit la réponse aux stimuli.

Toujours selon lui, nos réponses plus réflexives (celles propres au S2) sont le produit de l'activité des PT2. Ces derniers réagissent lentement et de façon contrôlée aux stimuli et ils fonctionnent grâce à une « application délibérée de règles » (p.ex. logiques ou linguistiques) (Kahneman & Frederik 2002, p. 50). Enfin, les PT2 demandent un effort plus grand que les PT1 et quelques-unes des étapes qui produisent la réponse aux stimuli sont accessibles à la conscience.

Kahneman soutient également que les deux systèmes cognitifs diffèrent par leurs contenus : seuls les PT1 peuvent avoir (et ont souvent) un contenu affectif, tandis que les contenus des PT2 sont toujours neutres en termes de valence affective. Comme nous avons vu dans l'introduction à ce mémoire, les intuitions sont souvent le résultat d'heuristiques, entre autres, l'heuristique de l'affect [*affective heuristic*] (Slovic et al. 2007). On pourrait dire que le contenu affectif qui, selon Kahneman, est typique des processus intuitifs correspond aux heuristiques affectives.

Il faut ajouter une dernière précision : en disant ceci, je ne veux pas dire que pour Kahneman les intuitions ont *toujours* un contenu affectif. En effet, il mentionne quelques autres types de contenu propre aux intuitions, les prototypes, par exemple, mais je ne m'intéresserai pas à cet aspect de sa théorie ici. Il suffit pour

nous de savoir que l'hypothèse selon laquelle seulement les PT1 peuvent avoir un contenu affectif (et donc qu'elles fonctionnent sur la base d'heuristiques affectives) sera endossée par Greene et par Haidt.

1.1.2 Comment écarter un problème des théories des systèmes duaux

Dans la sous-section précédente, j'ai mentionné quelques-unes des propriétés qui, selon Kahneman, caractérisent les PT1 et les PT2. Cela nous conduit à nous poser la question suivante : à quelle condition pouvons-nous affirmer qu'un processus cognitif donné tombe dans une des deux catégories ? La réponse à cette question n'est pas aussi simple qu'il peut le sembler à première vue parce qu'elle nécessite que nous interprétions les listes que dresse Kahneman.

Une interprétation possible de ces listes est celle que nous pouvons appeler *l'hypothèse des grappes nomologiques* [nomological clusters]. Selon cette lecture, nous pouvons concevoir les propriétés caractérisant chacun des deux types de processus comme constituant des grappes nomologiques : c'est-à-dire que ces propriétés seraient censées avoir une « forte tendance à se présenter ensemble » (Kelly, Stich, Haley, Eng & Fessler 2007, p. 119). Selon cette lecture, constater, par exemple, qu'un processus requiert peu d'effort cognitif, nous pousserait à lui assigner toutes les autres propriétés caractéristiques des PT1 (p. ex., l'automaticité, la vitesse, l'opacité, etc.) et inversement pour les PT2 (Samuels 2009, p. 140).

Cependant, il est problématique de concevoir les propriétés comme étant des grappes nomologiques, et cela puisqu'il est possible que certains processus aient des propriétés (non contradictoires) appartenant typiquement à chacune des deux sortes de processus. Par exemple, une opération cognitive peut demander peu

d'effort tout en étant sous notre contrôle (i.e. aller en vélo) (l'efficiencia étant une caractéristique des PT1 et la contrôlabilité une propriété typique des PT2 ; Gawronski 2014, p. 5-6 ; Samuels 2009, p. 140-141).⁸ Cette interprétation des listes de propriétés mises de l'avant par Kahneman ne pourrait donc pas tenir en compte de tels cas.

Une autre interprétation possible est celle de *l'hypothèse disjonctive*. Selon une telle lecture, nous pourrions concevoir les propriétés énumérées par Kahneman comme étant indépendantes les unes des autres. Il suffirait ainsi qu'un processus présente une seule des propriétés typiques des PT1 (ou des PT2) pour que nous l'identifions comme étant un PT1 (ou un PT2) (Gawronski 2014, p. 6).

Or, selon Gawronski (2014, p. 6) considérer un processus comme étant de type 1 ou de type 2 sur la base d'une seule propriété est tout aussi problématique. D'une part, cette interprétation pourrait nous conduire à considérer un processus comme étant un PT1 pour certaines de ses propriétés (p.ex. puisqu'il est inconscient), sans qu'il soit de type 1 sous d'autres aspects (p.ex. s'il n'est pas efficient). Selon Gawronski, cela entraînerait une certaine contradiction : un processus serait PT1 et non-PT1 en même temps. D'autre part, l'interprétation disjonctive fait en sorte

⁸ L'hypothèse des grappes nomologiques pose problème aussi parce qu'il est difficile, dans les faits, qu'une opération cognitive présente toutes les propriétés d'une des deux catégories. En d'autres mots, c'est trop demander à un processus cognitif de répondre à toutes les exigences nécessaires pour faire partie d'une des deux classes de processus. À ce propos, il est vrai que généralement les traits typiques des deux classes de processus sont mutuellement exclusifs. Conséquemment, si un processus n'est pas automatique il sera contrôlé. Cela veut dire que, bien qu'il ait toutes les autres caractéristiques des PT1, il sera tout de même à cheval entre les deux classes. Nous avons déjà illustré ces cas dans le corps du texte. Mais ce ne sont pas toutes les propriétés qui caractérisent les deux catégories de processus qui sont exclusives. Il se peut, par exemple, qu'un processus ne fonctionne pas de façon associative (qui est une caractéristique typique des PT1), cela ne veut pas dire qu'il opère en suivant des règles (qui est la façon d'opérer typique des PT2). Conséquemment, l'hypothèse des grappes nomologique nous empêcherait aussi de classer ce genre de cas dans une des deux catégories.

que nous pourrions catégoriser deux processus comme étant de type 1 sur la base de la possession de deux caractéristiques différentes, selon Gawronski. Par exemple, nous pourrions affirmer que les processus α et β sont tous les deux des PT1, mais pour des raisons différentes : α parce qu'il est rapide et β parce qu'il est opaque. Plusieurs processus feraient donc partie de la même catégorie, sans nécessairement avoir quelque chose en commun. Cela pourrait entraîner des « confusions à propos de résultats conceptuellement distincts qui seront désignés par le même terme [PT1 ou PT2] » (Gawronski 2014, p. 6).

Cette dernière objection ne me semble pas constituer un problème qui nous empêcherait d'accepter l'hypothèse disjonctive. De fait, contrairement à Gawronski, je crois qu'une telle interprétation présente un grand avantage : elle nous permet de classer les processus qui sont à cheval entre les deux catégories (dont l'hypothèse des grappes nomologiques ne pouvait pas tenir compte). Dans le cadre de cette hypothèse, nous pourrions penser que plus un processus a des caractéristiques typiques des PT1, plus il appartient à cette catégorie, tout en portant attention aux propriétés le caractérisant, mais qui ne sont pas de type 1. Pour éviter des confusions il faudrait tout simplement être précis et rendre compte des raisons pour lesquelles nous classons les processus dans l'une ou dans l'autre des deux catégories. Cette interprétation des listes dressées par Kahneman semble être plus inclusive et plus flexible que l'hypothèse des grappes nomologiques. Elle semble donc plus apte à rendre compte de la complexité et de la variété des processus qui sont mis en place par notre esprit. Pour cette raison, dans le cadre de ce mémoire, j'endorsse l'hypothèse disjonctive à l'égard de la distinction entre PT1 et PT2.

1.1.3 La collaboration entre les deux systèmes

Kahneman croit que typiquement le S1 et le S2 collaborent dans la formation d'une réponse aux stimuli. Plus précisément, il affirme que :

[Le] Système 1 propose rapidement des réponses intuitives à des problèmes de jugement au moment où ils apparaissent et le Système 2 contrôle [*monitors*] la qualité de ces propositions [*proposals*], qu'il peut *endosser*, *corriger* ou *ignorer* [*override*] (Kahneman & Frederik 2002, p. 51, emphase ajoutée).

Dans cet extrait, Kahneman explique que les S1 offrent rapidement une réponse au stimulus, et que le S2 peut intervenir de trois différentes façons sur la réponse intuitive du S1 : il peut l'endosser, la corriger ou tout simplement décider de ne pas en tenir compte (voir la figure 1.1.3). Les PT2 ont donc le rôle de « vérificateurs » des résultats des PT1, bien qu'ils puissent parfois négliger d'effectuer la correction.

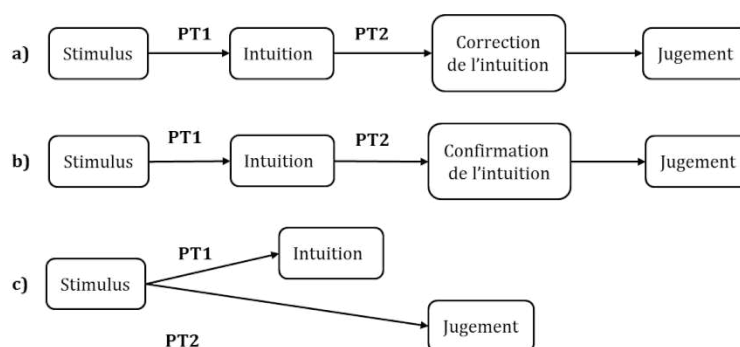


Fig. 1.1.3 Les trois types d'interaction entre S1 et S2 dans le modèle de Kahneman

Dans un premier cas (a), le S2 corrige la réponse intuitive offerte par le S1. Voici un exemple de ce cas de figure. Supposons qu'en sortant de l'université nous voyons une église dont les contours nous apparaissent plutôt flous. Selon

Kahneman, notre S1 est prédisposé à conclure que le bâtiment doit être loin de nous. En effet, selon cet auteur, le système intuitif évalue les distances en suivant cette règle : le moins les contours d'un objet sont nets, le plus cet objet est loin de nous. Cependant, les contours d'un objet nous paraissent flous aussi dans d'autres circonstances, notamment lorsqu'il y a du brouillard. Selon Kahneman, dans une telle situation, le S1 se trompe puisqu'il suggère que les objets aux contours flous sont plus loin qu'ils ne le sont réellement (Kahneman & Frederik 2002, p. 52). Cela explique pourquoi la présence du brouillard nous induit en erreur dans l'évaluation des distances. Or, notre S2 peut apprendre avec l'expérience à corriger notre jugement de distance en prenant en considération qu'il y a du brouillard. Malgré cela, selon Kahneman :

Bien que les gens soient capables de corriger consciemment leurs impressions de distance causées par la brume environnante, ils échouent généralement à le faire (Kahneman & Frederik 2002, p. 52).

Comme Kahneman et Frederik l'affirment dans cet extrait, savoir que le brouillard rend flou les contours des objets ne suffit pas pour que le S2 intervienne chaque fois pour réviser notre jugement de perception. Selon ces auteurs, l'intervention du S2 demande un effort que notre esprit n'est pas toujours prêt à faire ou qui demande de reconnaître la présence d'un facteur qui affecte notre évaluation. Cela nous donne un exemple du deuxième cas dans la figure (b) où le S1 donne une réponse erronée que le S2 néglige de corriger et qu'il confirme. Il peut par contre arriver aussi que le S2 confirme la réponse du S1 puisqu'il la considère correcte. Par exemple, dans le cas où il n'y a pas de brouillard et la distance du bâtiment est bien calculée par le S1.

En ce qui concerne le troisième cas de figure (c), Kahneman n'offre pas un exemple qui nous aide à distinguer les cas où le S2 ne tient pas compte de la réponse offerte par le S1 des cas où il la corrige. Nous pourrions par contre interpréter ce cas de figure de la façon suivante : on est toujours en train de

regarder l'église dans le brouillard, et le S1 nous porte à penser qu'elle est plus loin de ce qu'elle est de fait. Le S2 peut ignorer la réponse du S1 puisqu'il sait déjà quelle est la distance entre l'université et l'église (par exemple, puisque on l'a mesurée précédemment, ou qu'on l'a lu dans un texte etc.). Dans ce cas de figure le S2 ne tient tout simplement pas compte de l'intuition du S1. Par contre, dans le cas de figure (a), il tenait compte de la réponse intuitive en la corrigeant ensuite.

1.1.4 Flexibilité et apprentissage des deux systèmes

Jusqu'à maintenant, j'ai présenté les traits caractéristiques des deux types processus distingués par Kahneman et j'ai donné des exemples des phénomènes que sa théorie peut expliquer. Il faut en dernier lieu souligner que pour Kahneman un processus cognitif peut passer d'un système à l'autre. Par exemple, une tâche qui requiert initialement de l'effort cognitif, notamment jouer aux échecs, que nous pouvons considérer comme faisant partie du S2, peut devenir automatique avec le temps, et ainsi « rejoindre » les PT1 qui composent le S1. À ce propos, Kahneman affirme que :

[Les] opérations cognitives complexes migrent éventuellement du Système 2 au Système 1 au fur et à mesure que les compétences et les habiletés [*skills*] sont acquises. Une démonstration saisissante de l'intelligence du Système 1 est l'habileté des joueurs d'échecs à percevoir instantanément les forces et les faiblesses des positions d'échecs (Kahneman, D. & Frederick, S. 2002, p. 51).

Cet aspect du modèle de Kahneman justifiera l'argument que nous proposerons dans la suite de ce mémoire en faveur de la possibilité d'entraîner certains PT2, nous permettant ainsi de réviser nos jugements moraux afin de les rendre rapides et automatiques.

Nous avons vu jusqu'à maintenant que Kahneman distingue entre deux types de systèmes qui sont responsables du fonctionnement de l'esprit dans plusieurs de ses activités. Nous avons aussi montré comment ces deux processus interagissent afin de fournir une réponse aux stimuli et nous avons montré qu'il est possible, selon Kahneman, qu'un processus de type 2 puisse devenir de type 1 avec le temps et l'exercice. Ayant clarifié tout cela, nous pouvons maintenant présenter les modèles du jugement moral conçus par Haidt et par Greene, qui reposent sur cette théorie. J'expliquerai le modèle de Greene en premier dans la prochaine section.

1.2 Une solution possible

Comme je l'ai mentionné dans l'introduction, plusieurs recherches dans les dernières années indiquent que nos réponses affectives ont une grande influence sur nos jugements moraux. Les expériences conduites par Greene depuis 2001 ont confirmé de tels résultats. Sur cette base, Greene construit un modèle du jugement moral qui illustre comment l'influence de notre réaction émotionnelle sur le jugement moral change dépendamment de sa force et du temps que nous avons à disposition pour formuler le jugement. Étant donné que ce mémoire porte sur les méthodes qui peuvent être employées afin de diminuer l'influence des biais cognitifs sur les jugements moraux, nous allons porter notre attention à cet aspect de la théorie de Greene dans cette section, tout en montrant ses limites. Cela ouvrira la porte à une réponse alternative basée sur la théorie de Haidt, qui sera le sujet de la dernière section de ce chapitre.

1.2.1 Favoriser les raisonnements conséquentialistes

Selon Greene, la connaissance des processus qui sont responsables de la formulation de nos jugements moraux peut « altérer » notre attitude envers ces jugements (Greene 2001, p. 2017). Grosso modo, la connaissance fournie par la psychologie et les neurosciences au sujet du fonctionnement des mécanismes donnant lieu aux jugements moraux a un effet prophylactique sur ces derniers. Plus précisément, à la façon de Kahneman, mais dans le domaine moral, Greene soutient que nos jugements sont le fruit de l'activité combinée des intuitions et des raisonnements. Il affirme ensuite que dans certaines conditions (p. ex., lorsque nous jugeons hâtivement une situation), les jugements reflètent nos intuitions. Ces intuitions étant souvent tendancieuses, il soutient qu'il est préférable d'en douter. Il résume ainsi sa position :

La science peut faire progresser l'éthique en révélant le fonctionnement caché de nos jugements moraux, notamment ceux que nous faisons intuitivement. Une fois que leur fonctionnement interne est révélé, nous pouvons voir notre confiance diminuer envers certains de nos jugements ainsi que les théories éthiques qui sont (explicitement ou implicitement) basées sur eux (Greene 2014, p. 2).

Dans cet extrait Greene affirme que les recherches dans les domaines scientifiques peuvent justifier l'attitude sceptique qu'il suggère d'adopter envers nos jugements intuitifs. En outre, il fait allusion au fait que certaines théories éthiques soutiennent les jugements moraux découlant des intuitions. Plus précisément, il soutient que les jugements moraux déontologiques sont souvent générés par les réponses automatiques et émotives, alors que les jugements moraux conséquentialistes découlent généralement du raisonnement (Greene 2014, p. 7). Greene conclut que nous devrions non seulement « nous méfier » des intuitions, mais aussi qu'il faudrait favoriser une approche conséquentialiste dans le cas des dilemmes moraux (Greene 2014, p. 35). Ce faisant, Greene prend position à

faveur des jugements conséquentialistes, sans pourtant justifier ce choix explicitement.⁹

En conclusion, la méthode proposée par Greene consiste à douter de nos intuitions et réfléchir davantage sur l'action que nous sommes en train d'évaluer en tenant compte surtout de ses conséquences sur le bonheur général.¹⁰ Ce propos trouve son origine dans le modèle du jugement moral formulé par Greene et qui est caractérisé principalement par deux suppositions: a) nos jugements moraux découlent de l'activité de deux processus, dont les uns sont émotifs et les autres sont neutres; b) la réflexion, qui agit lentement, favorise la formulation de jugements moraux conséquentialistes tandis que les intuitions favorisent l'élaboration de jugements moraux déontologiques. Dans la prochaine section, je montrerai comment Greene vérifie ces deux hypothèses dans ses travaux empiriques.

1.2.2 *Trolleyologie*

Dans ses expériences, Greene commence par présenter aux participants des dilemmes comme les suivants :

⁹ Pour Greene, les jugements moraux conséquentialistes sont ceux que nous justifions en faisant appel au principe suivant : « une action est morale si elle maximise le bonheur, en tenant compte du fait qu'il n'y a personne dont le bonheur vaut plus que celui des autres » (Greene 2013, p. 170). Quant aux jugements moraux déontologiques, ils sont justifiés en faisant appel à ce qu'on considère comme étant des droits ou des devoirs des êtres humains.

¹⁰ Je crois que, pour Greene, cette méthode de correction du jugement moral peut être adoptée au moment où on formule le jugement (de façon synchronique) pour éviter de juger selon nos intuitions, mais elle peut aussi être employée pour corriger nos jugements intuitifs après que nous les avons formulés (de façon diachronique). Par souci de précision, il faut dire que Greene décrit seulement les cas où nous corrigeons notre jugement de façon synchronique, afin de raisonner de façon conséquentialiste.

Dilemme du Trolley : supposez que vous voyez cinq personnes ligotées à un chemin de fer et que vous voyez un train avançant dans leur direction. Vous avez la possibilité de tirer un levier pour changer la direction du train. Malheureusement, sur la direction alternative est ligoté quelqu'un d'autre. Est-il acceptable pour vous de tirer le levier, et ainsi, de sacrifier une personne pour en sauver cinq ?

Alternativement, imaginez faire face à la situation suivante :

Dilemme du Footbridge : Vous assistez à la même scène d'une passerelle qui est en haut de la voie ferrée. À côté de vous, il y a un homme corpulent. Vous savez qu'en le poussant sur le chemin de fer vous pourrez sauver les cinq personnes en bas de la passerelle. Est-il acceptable pour vous de le pousser ?

Selon les résultats des recherches effectuées par Greene (Greene 2001, p. 2105), la majorité des participants jugent qu'il est acceptable de tirer le levier pour faire changer la direction au train, mais qu'il est inacceptable de pousser la personne corpulente en bas de la passerelle. En effet, très peu des participants considèrent comme acceptable cette dernière option.¹¹

Comprendre pourquoi nous répondons différemment aux deux situations a longtemps occupé les chercheurs en psychologie morale (Greene 2001, p. 2106). Pour sa part, Greene formule l'hypothèse suivante afin de rendre compte des résultats obtenus : il soutient que certains dilemmes causent une réponse émotive

¹¹ Greene ne mentionne pas le pourcentage exact.

plus forte que d'autres, et que cette différence influence les jugements moraux (Greene 2001, p. 2106). On peut donc supposer que Greene serait d'accord pour affirmer que nos jugements moraux découlent dans ces cas d'une heuristique de l'affect. Il affirme à ce propos :

Nous pouvons soutenir que, d'un point de vue psychologique, la différence cruciale entre le problème du Dilemme du *Trolley* et le Dilemme du *Footbridge* repose dans la tendance qu'a ce dernier à activer [*engage*] les émotions des gens par opposition au premier (Greene 2001, p. 2106).

De façon plus générale, Greene croit que les dilemmes qui impliquent une violation morale personnelle, à savoir, ceux semblables au dilemme du Footbridge, « engagent les processus émotionnels » davantage que ceux qui impliquent une violation morale impersonnelle comme le dilemme du Trolley. Pour comprendre ces affirmations, il faut faire un pas en arrière et expliquer ce que Greene entend par « violation morale personnelle ».

Greene affirme qu'une action est une violation morale personnelle si elle respecte les trois critères suivants :

Premièrement, il doit être probable que la violation cause un tort corporel sérieux [*serious bodily harm*]. Deuxièmement, ce tort doit arriver à une personne ou à un ensemble de personnes. Troisièmement, le tort ne doit pas être le résultat d'une *déflexion d'une menace existante vers un tiers* (Greene 2004, p. 389, emphase ajoutée).

En d'autres mots, une violation morale est personnelle s'il est très probable que sa performance nuise physiquement à quelqu'un, sans que le mal provoqué par l'action soit le résultat du fait que cette même action permet de sauver une autre personne d'un danger. Ensuite, Greene ajoute :

L'on peut concevoir ces trois critères en termes de « MOI FAIS MAL TOI » [*ME HURT YOU*]. Le (i) critère 'FAIS MAL' désigne les violations nuisibles [*harmful*] du genre le plus primitif (p. ex., une agression plutôt qu'un délit d'initié) alors que (ii) le critère "TOI" assure que la victime est nettement représentée comme un individu. Finalement, (iii) *la condition "MOI" rend une notion 'd'agentivité' demandant que l'action surgisse de manière directe de la volonté de l'agent, qu'il en soit l'auteur plutôt que simplement l'exécutant* (Greene 2004, p. 389, emphase et énumération ajoutées).

Ce deuxième extrait présente deux problèmes d'interprétation. D'une part, à première vue, dans ce dernier passage Greene semble modifier le troisième critère dans la définition des violations morales personnelles. En effet, ici le troisième critère implique que *l'agent soit l'auteur et non pas simplement l'exécuteur de l'action*. Par contre, dans le passage précédent, le troisième critère requérait que *le mal qu'on cause à autrui ne soit pas le résultat d'une action faite pour éloigner un danger d'une autre personne*. D'autre part, Greene affirme, dans ce deuxième passage, qu'une violation morale est personnelle si l'action découle directement de la volonté de l'agent, à savoir, si l'agent en est l'auteur. Pourtant cette définition ne nous permet pas de comprendre pourquoi le dilemme du Footbridge serait une violation morale personnelle contrairement au dilemme du Trolley. En effet, dans tous les deux de ces cas, l'action découle de la volonté de l'agent (tirer le levier vs pousser la personne corpulente en bas de la passerelle).

Toutefois, une analyse plus détaillée de ces deux passages nous permet de résoudre ce problème d'interprétation. En effet, Greene affirme dans une note à la fin de son article, que :

Lorsqu'un tort est produit par le biais de la déflexion d'une menace existante, *l'agent est simplement "l'exécuteur"* [*has merely edited*] *et non pas "l'auteur" du tort qui en résulte*, et donc, sa considération [*contemplation*] est moins engageante émotionnellement (Greene 2001, p. 2107, emphase ajoutée).

Dans cette note, Greene explique quel est le lien entre les deux formulations qu'il donne du troisième critère. En d'autres mots, pour lui, si une action est mise en place pour détourner un danger qui menace autrui (par exemple, celle de tirer le

levier dans le cas du dilemme du Trolley) alors l'agent n'est pas l'auteur, mais l'exécuteur du mal qui peut éventuellement s'en suivre (notamment la mort de l'homme ligoté sur l'autre chemin de fer). Ce passage nous fait aussi comprendre que la confusion dans l'interprétation du deuxième passage dérive du fait qu'ici Greene emploie le terme « action » pour désigner *le résultat nocif de l'action* et non l'action en elle-même, comme il le précise dans la note qu'on vient de citer. En d'autres mots, pour lui, une violation morale est personnelle si le résultat nocif de l'action qu'on décide d'accomplir est voulu par l'agent, de sorte que l'agent en soit l'auteur et non seulement le réalisateur. Par contre, si le mal découle non pas de la volonté de l'agent, mais il est plutôt une conséquence non désirée de l'action en elle-même, comme une conséquence non souhaitée, alors la violation morale est impersonnelle. Conséquemment, la violation morale du dilemme du Trolley n'est pas personnelle puisque le mal qui suit du fait qu'on tire le levier ne « surgit pas directement de la volonté de l'agent » (Greene 2004, p. 389). Dans ce cas, l'agent est simplement l'exécuteur et non pas l'auteur du mal infligé à autrui. Conséquemment, dans le cas du dilemme du Trolley, il s'agit d'une violation morale impersonnelle.

Nous pouvons maintenant rediriger notre attention sur l'analyse des expériences par lesquelles Greene (2001) chercha à établir que la différence des jugements sur le dilemme du Trolley et le dilemme du Footbridge découlait des diverses réactions émotives que ces deux types de dilemmes provoquent.

Afin de valider cette première hypothèse, Greene a examiné les résultats d'imagerie par résonance magnétique fonctionnelle des participants répondant aux deux dilemmes. Tel que prévu, il a constaté que les régions cérébrales correspondantes aux émotions s'activent davantage lorsque les participants répondent au dilemme du Footbridge. Ces données supportent son hypothèse plus générale selon laquelle les violations morales personnelles (p.ex. le dilemme du

Footbridge) causent une réaction émotionnelle plus intense que les violations morales impersonnelles (p.ex. le dilemme du Trolley). Conséquemment, l'hypothèse de Greene selon laquelle nos jugements peuvent découler ou bien d'un processus cognitif chargé émotionnellement ou bien d'un processus cognitif plus « neutre » est supportée par ces évidences.

Dans des recherches subséquentes (Greene 2004 et 2012), Greene chercha à montrer que la réflexion peut modifier le jugement moral. Pour ce faire, il définit tout d'abord quelles sont pour lui les deux caractéristiques principales des processus réflexifs. Il affirme ainsi que les processus cognitifs plus rationnels sont sensibles à la force d'un argument et qu'ils demandent plus de temps pour être mis en place. En effet, il affirme :

Un processus de jugement qui échoue à distinguer entre arguments forts et arguments faibles pourrait difficilement être considérés comme étant une forme de "raisonnement", peu importe sa durée (Greene 2012, p. 169).

Il dirigea ensuite son attention sur les actions que nous considérons comme des violations morales personnelles et sur les cas dans lesquelles nous estimons ces mêmes actions comme étant acceptables. En effet, comme nous avons vu, les cas de violations morales personnelles déclenchent une forte réponse émotionnelle (intuitive) qui détermine souvent le jugement. Le but de Greene est de montrer qu'avec la réflexion nous pouvons changer ce jugement intuitif. Par exemple, il présenta le dilemme suivant aux participants à ses expériences :

John est le capitaine d'un sous-marin militaire voyageant sous un énorme iceberg. Une explosion à bord du vaisseau fait en sorte que celui-ci perd la majeure partie de ses réserves d'oxygène ; un marin est blessé et perd son sang rapidement. Le marin blessé va succomber de ses blessures, peu importe ce qui se produira. L'oxygène n'est pas suffisant pour que l'équipage au complet se rende jusqu'à la surface. La seule manière de sauver les autres membres de l'équipage est que John tue le marin

blessé de sorte qu'il restera tout juste assez d'oxygène pour que le reste de l'équipage puisse survivre (Greene 2012, p. 165).

Dans ce cas, le jugement conséquentialiste serait de tuer le marin pour sauver le reste de l'équipage. Suite à la lecture du dilemme, Greene présenta à certains participants un argument très convaincant en faveur du jugement conséquentialiste, alors qu'à d'autres participants il présenta un argument résolument moins convaincant en faveur du même jugement. L'expérience était conçue de manière telle que tous les participants pouvaient prendre leur temps pour lire l'argument, mais qu'environ la moitié de ceux-ci étaient encouragés à y réfléchir davantage (soit pour deux minutes de plus). Les résultats de cette expérience selon Greene montrent que :

La réflexion n'augmente pas seulement l'effet de la force de l'argument. Il n'y avait pas d'effet de la force de l'argument lorsque la réflexion n'était pas encouragée [...] nous avons découvert qu'un argument fort était plus convaincant qu'un argument faible, mais seulement lorsque les sujets étaient encouragés à réfléchir (Greene 2012, p. 171-172).

En d'autres mots, Greene remarque qu'une variation du temps que nous accordons à la réflexion ainsi qu'un encouragement à réfléchir, entraînent souvent un changement dans les jugements moraux. Il montre ainsi, comme prévu, qu'une variation du temps que nous prenons pour répondre aux dilemmes qui provoquent une forte réponse émotionnelle peut entraîner un changement dans nos jugements. Plus précisément, il montre que les jugements conséquentialistes sont favorisés par la réflexion.

En conclusion, les expériences de Greene (2001, 2004 et 2012) permettent de soutenir les deux hypothèses suivantes : a) nos jugements moraux déontologiques découlent des intuitions chargées émotionnellement b) la réflexion, qui agit lentement, favorise la formulation de jugements moraux conséquentialistes. Ces deux

hypothèses caractérisent le modèle de jugement moral que Greene propose, modèle dont nous allons maintenant décrire les aspects généraux.

1.2.3 Le modèle de Greene

En fondant sa théorie en large partie sur les expériences que nous venons de mentionner, Greene soutient que l'esprit fonctionne grâce à l'activité de deux systèmes cognitifs, et il endosse explicitement la théorie de Kahneman à ce propos. Il affirme ainsi que :

Le cerveau humain a le même design. Premièrement, nous avons une variété de réglages [*settings*] automatiques—des réflexes et des intuitions qui guident notre comportement, dont plusieurs sont (a) émotionnels. Nous pouvons être conscients de ces réponses émotionnelles, mais nous ne sommes généralement (b) pas conscients des processus qui les déclenchent. Nous comptons sur nos réglages automatiques la plupart du temps, ceux-ci nous sont utiles en règle générale. Nos cerveaux ont aussi un mode manuel. [...] Les opérations de ce système sont typiquement (1) conscientes, (2) sont perçues comme volontaires, et souvent, comme demandant de (3) l'effort. [...] En bref, le mode de pensée manuel est le type de pensée que considérons comme étant la “pensée” (Greene 2014, p. 3, énumération ajoutée).

Pour Greene, les intuitions sont (a) émotivement chargées et (b) les processus qui les produisent ne sont pas accessibles à la conscience. J'aimerais souligner le fait que comme pour Greene les intuitions semblent avoir toujours un contenu affectif lorsqu'elles déterminent le contenu de nos jugements moraux, cet auteur serait probablement d'accord dans l'affirmer que les intuitions morales sont le fruit d'une heuristique de l'affect telle que présentée par Slovic (et al. 2007). Par contre, les processus qui constituent le système rationnel sont (1) conscients (2) délibérés et (3) ils demandent de l'effort. Selon Greene celui des deux processus qui influencera le plus le contenu du jugement dépend de deux variables : la force de l'émotion qui surgit automatiquement suite au stimulus et le temps de réponse dont nous disposons.

D'un côté, il soutient que si l'émotion déclenchée par le stimulus est forte, le temps que nous avons pour répondre sera déterminant. En effet, lorsque nous n'avons pas le temps de réfléchir davantage sur la question, la réponse émotive des PT1 sera la réponse finale. Conséquemment, le jugement sera déontologique. Par contre, si nous avons plus de temps pour réfléchir, les PT2 pourront intervenir et il sera possible de substituer la réponse émotivement chargée, avec une réponse plus « froide ». En raison de l'intervention de la réflexion, le jugement intuitif sera remplacé par un jugement conséquentialiste. De cette manière, pour la plupart des gens, décider de pousser un homme en bas d'une passerelle, comme dans le dilemme du Footbridge, est inacceptable puisque cette situation provoque en nous une forte réponse émotive qui empêche la raison de prendre le contrôle. Pour que la réflexion puisse intervenir, il faut donner aux participants plus de temps pour répondre (figure 1.2.3).

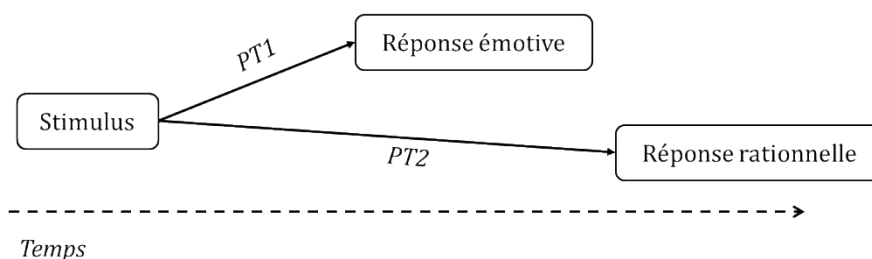


Fig. 1.2.3 L'activité des PT1 et des PT2 dans la formulation des jugements moraux en relation au temps

De l'autre côté, la réponse émotive aux stimuli pourrait être faible, comme dans le cas du dilemme du Trolley. Étant donné son intérêt pour les réactions aux violations morales personnelles qui déclenchent une forte réponse émotive, Greene néglige d'expliquer en détail ce qui se passe lorsque la réponse émotive au stimulus n'est pas très forte. Quel processus cognitif détermine que la plupart des participants considèrent acceptable de tirer le levier en réponse au dilemme du Trolley ? Greene n'offre pas de réponse claire à cette question. En effet, les résultats de ses recherches montrent que les participants répondent vite à ce type

de dilemme avec un jugement conséquentialiste, sans que ce dernier découle d'une forte réponse émotive. De toute façon, cette éventualité n'est pas intéressante de son point de vue, puisque dans le cas où la réaction émotive est faible, le jugement est conséquentialiste. Étant donné que, comme nous avons mentionné, pour Greene les jugements conséquentialistes sont corrects, ils ne nécessitent pas d'être corrigés.

1.2.4 Les limites de la méthode proposée par Greene

Selon Greene, notre esprit a besoin de temps et d'effort pour mettre en place des processus délibérés et rationnels. Conséquemment, une tactique que nous pouvons adopter afin de diminuer l'influence que les intuitions émotivement chargées (et donc l'heuristique de l'affect) ont sur la formulation des jugements moraux est très simple : nous devrions douter de nos jugements et prendre du temps afin de les corriger en favorisant les raisonnements conséquentialistes. Une telle suggestion présente trois problèmes à mon avis.

Premièrement, cela requiert que nous ayons conscience du fait que certains biais cognitifs peuvent influencer nos jugements moraux. Deuxièmement, cela implique que les mécanismes de contrôle et de correction des jugements peuvent intervenir à chaque occurrence. De plus, Greene néglige d'expliquer la nature de la motivation qui nous pousserait à corriger nos jugements chaque fois. Troisièmement, la suggestion de Greene de réfléchir surtout aux conséquences des actions afin de bien juger est problématique puisqu'elle n'est valide que pour un groupe restreint de jugements moraux. Je vais passer le prochain paragraphe à l'explication de cette dernière critique, puisqu'elle me semble la moins évidente.

Comme nous l'avons mentionné, pour Greene il faudrait prêter beaucoup d'attention aux situations qui déclenchent une forte réaction émotionnelle, puisque celle-ci fait obstacle au raisonnement conscient. Pourtant, dans sa théorie, il ne mentionne que des cas de violation morale personnelle (dilemme du Footbridge) comme exemples de situations qui déclenchent une forte réponse émotionnelle. Cela implique que de fortes émotions seraient déclenchées seulement dans les cas où le bien-être de quelqu'un d'autre est mis en jeu.

Toutefois, dans la littérature sur les jugements moraux, il y a de plus en plus d'articles qui montrent que des circonstances autres que les violations morales personnelles peuvent générer des réponses intuitives qui déterminent nos jugements moraux. Notamment, des émotions comme le dégoût influencent les jugements, même dans des situations où personne ne subit un tort (p.ex. Haidt 1993 ; Wheatley & Haidt 2005). Voici un exemple où l'émotion négative est déclenchée par un comportement qui implique un manque de respect, et qui détermine notre jugement sur l'action :

Une femme sur son lit de mort demande à son fils de lui promettre de visiter sa tombe chaque semaine. Le fils a beaucoup aimé sa mère, alors il lui promet de visiter sa tombe chaque semaine. Mais lorsque la mère meurt, il ne tient pas sa promesse, car il est très occupé (Haidt 1993, p. 617).

En outre, Greene affirme qu'une approche conséquentialiste devrait être favorisée afin de répondre adéquatement aux problèmes moraux. Donc, à son avis les réflexions qui doivent s'occuper de réviser les jugements hâtifs qu'on formule suite à une forte réponse émotionnelle évaluent les actions en considérant leurs conséquences vis-à-vis de la maximisation du bonheur. Cependant, si nous prenons comme exemple la situation que je viens de décrire, nous n'avons pas les éléments nécessaires pour évaluer les conséquences de l'action en question sur la maximisation du bonheur. En effet, nous ne savons pas si quelqu'un est blessé à cause du fait que le fils n'honore pas la promesse faite à sa mère. De plus, même

si le résultat de l'action en question est nocif, la violation morale demeure impersonnelle puisque le mal qui en suit n'est pas voulu par l'agent (en tout cas, la vignette laisse suggérer que l'agent n'a pas brisé sa promesse intentionnellement). Dans ce cas, un raisonnement conséquentialiste serait tout à fait inutile.

En conclusion, je crois que la solution proposée par Greene pour corriger les jugements découlant d'une forte réponse affective peut être adoptée sur une partie très restreinte de la panoplie de situations qui peuvent susciter des réponses émotives influençant nos jugements moraux. De fait, nous pourrions suivre sa suggestion seulement dans les situations de violation morale personnelle. Cette limite de la méthode de correction des jugements moraux proposée par Greene dérive du fait qu'il élabore son modèle explicatif du jugement moral en partant d'expériences où les participants peuvent formuler seulement deux types de jugements moraux : les uns déontologiques (qui se trouvent être intuitifs) et les autres conséquentialistes (qui se trouvent être réflexifs).

Dans la section suivante je présenterai le modèle socio-intuitionniste du psychologue Jonathan Haidt, qui peut être conçu comme étant une expansion de celui de Greene. En effet, nous verrons que pour Haidt nos jugements moraux découlent également de l'activité de deux processus, dont les uns sont émotifs et rapides tandis que les autres sont neutres et lents. Toutefois, à la différence de Greene, Haidt tient compte des divers types de stimuli qui génèrent des fortes réponses affectives en nous et qui peuvent donc provoquer des jugements intuitifs. D'ailleurs, Haidt ne caractérise pas les jugements intuitifs comme étant déontologiques et les jugements réflexifs comme des conséquentialistes. Enfin, Haidt ne semble avoir une attitude plus positive envers les intuitions morales : à la

différence de Greene, il ne nous encourage pas à adopter une attitude sceptique envers nos jugements moraux intuitifs. En conséquence de ce qui précède, le modèle de Haidt a une portée explicative plus importante que celui de Greene, c'est-à-dire qu'il peut permettre d'expliquer l'influence que les intuitions ont sur nos jugements dans plusieurs situations différentes. Pour cette raison, selon moi, ce modèle du jugement moral constitue un meilleur point de départ afin de décrire les opérations mentales qui nous permettent de formuler des jugements moraux. Nous verrons s'il est possible de proposer une méthode pour diminuer l'influence des intuitions chargées émotionnellement sur nos jugements moraux à partir de ce modèle. Je vais donc maintenant présenter de façon plus détaillée le modèle de Haidt afin d'en montrer les avantages et les limites.

1.3 Le conducteur et l'éléphant

Cette section du chapitre a pour but d'examiner en détail la théorie « socio-intuitionniste » de Haidt (2001). Ce faisant, je vais me pencher plus en détail sur l'analyse des opérations cognitives qui, selon lui, nous permettent de modifier nos jugements moraux.

1.3.1 Le modèle socio-intuitionniste de Haidt

Haidt nomme son modèle du jugement moral « socio-intuitionniste » en soulignant ainsi ses deux aspects principaux. Son modèle est « intuitionniste », car il implique que nos jugements moraux sont souvent déterminés par nos intuitions plutôt que par le raisonnement, et « social » puisque ces intuitions sont influencées par le contexte social.

Comme illustré dans le tableau 1.3.1, pour Haidt les intuitions sont (a) rapides, (b) sans effort et (c) automatiques. De plus, (d) seul le résultat de leur fonctionnement est accessible à la conscience et (e) leur contenu est affectif. Au contraire, le raisonnement a lieu (1) plus lentement (2) en demandant de l'effort cognitif, et (3) certaines de ses étapes sont accessibles à la conscience (Haidt 2001, p. 818). Comme nous pouvons le remarquer, Haidt accepte la distinction entre deux types de processus cognitifs offerte par Kahneman (Haidt 2012a, p. 869).

PT1 - Intuitions	PT2 - Délibérations
(a) Rapide	(1) Lent
(b) Sans effort	(2) Avec effort
(c) Automatique	Contrôlé (?)
(d) Inconscient	(3) Conscient
(e) Affectif	Neutre (?)

Tab. 1.3.1 Les caractéristiques des processus duaux impliqués dans l'élaboration des jugements moraux (Jonathan Haidt)

J'aimerais maintenant attirer l'attention sur le type de contenu qui est propre aux deux types de processus. Pour Haidt les affections qui constituent le contenu de nos intuitions sont des « apparitions » [*flashes*] de sentiments positifs ou négatifs (Haidt 2012b, p. 65).

De plus, il affirme que nos intuitions morales « comprennent » une émotion morale (Haidt 2012b, p. 58). La relation entre ces deux types de contenus (l'émotion et l'affection) des intuitions est la suivante : une émotion inclut toujours une affection, mais une affection ne peut pas constituer une émotion par elle-même puisqu'elle est « trop fugace » [*too fleeting*] (Haidt 2012b, p. 65).¹² Donc, les intuitions comprennent toujours des émotions morales et ces dernières incluent toujours un sentiment positif ou négatif. Au contraire, les délibérations

¹² Sur la possibilité que les émotions soient uniquement « cognitives » voir Prinz, 2009.

sont dépourvues de charge affective. J'aimerais faire noter au lecteur que Haidt, comme Greene, considère que nos jugements intuitifs contiennent une affection. Cela nous autorise à croire que Haidt serait lui aussi d'accord avec Slovic (et al. 2007) pour qui nos intuitions sont souvent le résultat d'une heuristique de l'affect qui peut parfois nous induire en erreur.

En outre, comme je le montrerai plus tard, le fait que les intuitions comprennent une affection dans le modèle de Haidt implique que lorsqu'un jugement est déterminé par les PT1, celui-ci motive à l'action, contrairement aux jugements qui découlent des PT2. Cela dit, nous pouvons maintenant examiner plus en détail quels sont les rôles que ces deux processus jouent selon Haidt dans l'élaboration des jugements moraux.

1.3.1.1 L'éléphant et le conducteur

Dans son livre *The Happiness Hypothesis* (2006), Haidt affirme que la raison (PT2) est pour les intuitions (PT1) ce qu'un conducteur d'éléphant est pour son éléphant. Il affirme que le conducteur est souvent porté par l'éléphant et qu'il est rarement libre de le guider dans l'élaboration des jugements moraux. Ainsi, selon Haidt, la première réponse à un « stimulus déclencheur » [*eliciting stimulus*] est une intuition (PT1) chargée émotionnellement (flèche1) (Haidt 2001, p. 819). Par la suite l'intuition « cause » le jugement moral (flèche2) et finalement des processus rationnels entrent en jeu lorsque nous devons justifier ce jugement à autrui (flèche3) (Haidt, 2001, p. 819). Il dit à ce propos que :

[L]e jugement moral est causé par des intuitions morales rapides, et il est suivi (si nécessaire) d'un raisonnement moral lent et ex post facto. (Haidt 2001, p. 814).

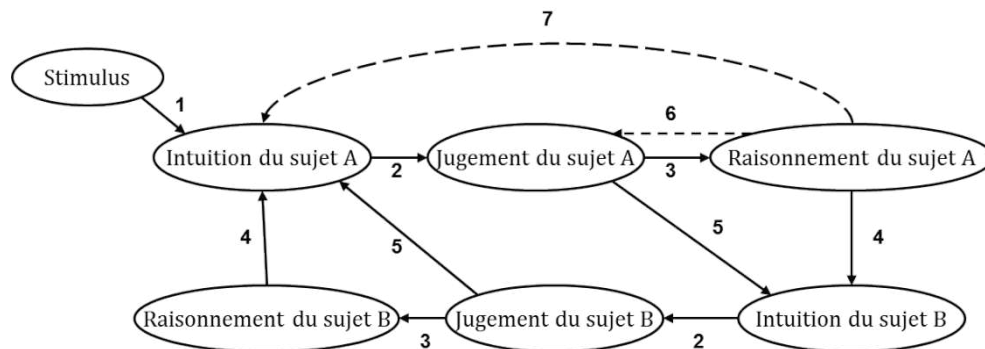


Fig. 1.3.1 Traduction du schéma du modèle socio-intuitionniste (Haidt 2012, p. 55)

1.3.1.2 Le conducteur d'autrui guide notre éléphant

Les flèches 4 et 5 dans la figure 5 illustrent l'influence que la vie sociale peut exercer sur nos intuitions. Selon Haidt, le raisonnement persuasif (flèches 4) des autres (sujet B) peut provoquer une nouvelle intuition en nous (sujet A), même si cela se produit rarement. Au contraire, il croit que (flèches 5) venir simplement à connaître les jugements des autres (sujet B) suscite facilement en nous (sujet A) de nouvelles intuitions (Haidt 2001, p.819). Passons maintenant à l'analyse des deux dernières flèches indiquant les cas où notre raison peut, par elle-même, modifier les jugements moraux.

1.3.1.3 Notre conducteur guide notre éléphant

Les flèches 6 et 7 dans la figure 1.3.1 sont pointillées afin d'indiquer que la réflexion (PT2) peut déterminer nos jugements et qu'elle peut, théoriquement, avoir une influence sur nos intuitions, mais qu'elle intervient « rarement dans la pratique » (Haidt 2012b, p. 54). Plus précisément, selon Haidt, deux types de réflexion modifient nos jugements : l'un directement et l'autre indirectement.

Dans le premier cas, le jugement est directement révisé par le raisonnement logique (flèche 6); alors que dans le deuxième cas, le changement de perspective (flèche 7) entraîne un nouveau jugement indirectement par le biais d'une nouvelle intuition. Les processus cognitifs concernés ici feront l'objet de l'essentiel de ma discussion du modèle de Haidt puisqu'ils sont responsables de la correction des jugements moraux. Dans la suite de ce texte, j'argumenterai afin de montrer qu'utiliser le changement de perspective (flèche 7) est plus avantageux qu'avoir recours au raisonnement logique (flèche 6) pour la révision des jugements moraux.

Jusqu'à maintenant, nous avons vu que selon le modèle socio-intuitionniste les jugements moraux entraînent deux types de processus, les uns intuitifs (PT1) et les autres délibératifs (PT2) et nous avons brièvement énuméré leurs caractéristiques respectives. Ensuite, nous avons vu comment ces deux types de processus entrent en jeu dans l'élaboration des jugements moraux. À la lumière de la description que je viens de donner du modèle socio-intuitionniste, il semblerait que les jugements moraux découlent souvent des PT1 et que les PT2 entrent plus souvent en jeu pour justifier les jugements moraux que pour leur élaboration. Toutefois, j'estime qu'en examinant en détail le modèle de Haidt nous verrons qu'il est possible que les PT2 sont impliqués aussi dans la formulation des jugements moraux dont le contenu est déterminé par les PT1. Nous montrerons aussi dans la prochaine sous-section que cette interprétation serait aussi plus cohérente avec la théorie de Kahneman à laquelle le modèle de Haidt est inspiré.

1.3.2 Les PT2 : non seulement des réflexions *post-hoc*

Haidt est clair lorsqu'il affirme que le raisonnement logique (flèche 6) et le raisonnement stratégique (Haidt 2012a, p. 869) suivent l'élaboration des jugements moraux. Pour Haidt, ces deux types de « raisonnement » sont des PT2 qui interviennent après que les PT1 ont déterminé les jugements moraux. Toutefois, cela n'exclut pas que d'autres PT2 puissent jouer un rôle dans l'élaboration des jugements moraux.

1.3.2.1 Le raisonnement déductif suit le jugement moral : le cas de la perplexité morale

Selon la reconstruction que Haidt fait de l'histoire des théories psychologiques morales, avant 2001, le rationalisme était la théorie dominante. Selon Haidt, les tenants du rationalisme attribuent un rôle fondamental aux processus rationnels dans l'élaboration des jugements moraux. Or, Haidt ne dit pas clairement quelle sorte de processus rationnel serait à l'origine des jugements moraux selon les rationalistes. Saunders (2009, p. 335) comble cette lacune en affirmant que pour les rationalistes les jugements moraux découlent des « déductions à partir des principes moraux qu'on détient consciemment ». Par exemple, ils pourraient affirmer que nous jugeons en suivant ce raisonnement :

Nuire à quelqu'un est mal ;

L'action de Sam a nui à Marie ;

Donc, l'action de Sam est mal.

Comme nous avons illustré, Haidt estime que, dans la majorité des cas, ce type de raisonnement déductif suit le jugement moral et il sert à justifier ce jugement. Haidt justifie une telle affirmation par la considération suivante : si les raisonnements déductifs étaient la cause de nos jugements, il devrait être simple pour nous de modifier ceux-ci lorsqu'on reconnaît qu'ils sont erronés ou mal fondés. La vie de tous les jours nous montre qu'au contraire nous éprouvons une difficulté à reformuler nos jugements pendant une discussion, même si nous avouons que les arguments de notre opposant sont valides. Nous avons également du mal à les modifier lorsque nous découvrons que nos raisons pour les poser ne sont pas valides. Ce phénomène est connu sous le nom de « perplexité morale » [*moral dumbfounding*]. Lorsque nous sommes moralement perplexes, nous ne modifions pas notre jugement malgré le fait que nous sommes incapables de lui donner une justification valide (Haidt 2001, p. 817).

Pour renforcer sa critique du rationalisme, Haidt rend compte de nombreuses expériences suscitant de la perplexité morale. L'une d'entre elles consistait à poser la question suivante aux participants :

Julie et Mark sont frère et sœur. Ils sont en train de voyager ensemble en France pendant leurs vacances suite à l'année universitaire. Une nuit, ils restent seuls dans un chalet près de la plage. Ils décident qu'il serait intéressant et amusant d'essayer de faire l'amour. Au minimum, cela serait une nouvelle expérience pour chacun d'entre eux. Julie prenait déjà la pilule contraceptive, mais Mark utilise un préservatif tout de même, afin d'être certain. Ils prennent plaisir tous les deux à faire l'amour, mais décident de ne plus le refaire à nouveau. Ils gardent cette nuit comme leur secret spécial, qui les fera se sentir plus proche l'un de l'autre. Que pensez-vous de tout cela, était-il OK pour eux de faire l'amour ? (Haidt 2001, p. 814).

Bien que la majorité des participants aient répondu que le comportement des deux protagonistes était moralement mauvais, ceux-ci étaient incapables de trouver des arguments valides pour justifier cette réponse. À titre d'exemple, certains justifiaient leur réponse en affirmant que les protagonistes pourraient avoir des enfants avec des problèmes de santé. Leur jugement restait identique (c'est-à-dire que ce que faisaient les protagonistes était mal) même après s'être fait rappeler

que les jeunes en question avaient utilisé deux moyens de contraception. Les participants cherchaient donc des arguments pour justifier leur jugement moral, plutôt que de réviser celui-ci. Selon Haidt, les théories rationalistes des jugements moraux ne pourraient pas expliquer comment il est possible de ne pas savoir pourquoi nous jugeons inacceptable le comportement de Julie et Mark. Il élabore donc un modèle de jugement moral qui rend compte de ce phénomène : le modèle socio-intuitionniste. Comme nous l'avons dit plus haut, dans ce modèle le raisonnement déductif et stratégique formule des justifications pour les jugements moraux qui ont été déterminés par les intuitions. Ils sont donc toujours *post-hoc*. Ainsi, les processus rationnels, qui selon les rationalistes causaient les jugements moraux, sont plutôt pour Haidt des processus de justification du jugement.

1.3.2.2 Les PT1 sont-ils suffisants pour élaborer un jugement moral ?

Pour le modèle socio-intuitionniste, un « jugement moral apparaît à la conscience automatiquement et sans effort comme le résultat des intuitions morales » (Haidt 2001, p. 819). Nous pourrions donc penser que pour Haidt l'intuition et l'émotion qui en constitue le contenu sont suffisantes afin d'élaborer un jugement moral. Cette idée me semble être mitigée par les affirmations de Haidt dans l'extrait suivant, où il maintient que :

Les règles et les jugements moraux « doivent porter sur les intérêts ou le bien-être, de la société en entière, ou minimalement sur ceux de personnes qui ne sont ni le juge, ni l'agent » (Gewirth 1984, p. 978). Cette seconde approche est plus prometteuse pour le travail psychologique, puisqu'elle ne lie pas la moralité au langage, permettant de la sorte des discussions sur *les origines des émotions morales* chez les animaux pré-linguistiques et chez les enfants (Haidt 2003, p. 853, emphase ajoutée).

Comme nous l'avons vu précédemment, les émotions morales constituent pour Haidt le contenu des intuitions, mais dans cet extrait il se montre seulement ouvert

à l'éventualité que les émotions morales trouvent leur *origine* chez les animaux et les enfants pré-linguistiques. En somme, avec ces affirmations Haidt me semble plutôt soutenir que notre capacité morale trouve son origine dans les émotions morales, à savoir, dans les intuitions et non pas qu'elle dépend *uniquement* de ces dernières.

Ailleurs il affirme que, bien avant que la raison soit bien développée, les enfants formulent quelque chose « comme des jugements moraux » [*something like moral judgments*] (Haidt 2012b, p.75). Cette citation laisse aussi penser que pour lui les émotions morales ne sont pas tout à fait suffisantes pour donner lieu à un « vrai » jugement moral et que les PT2 (que pour Haidt sont responsables des raisonnements) jouent peut-être un rôle important dans l'élaboration d'une « vraie » évaluation morale. En outre, il affirme que :

Plutôt que de chercher une définition formelle qui énumère les conditions nécessaires et suffisantes du jugement moral, cet article prend une approche plus empirique, commençant avec des faits comportementaux propres aux êtres humains : que dans chaque société (a) *les personnes parlent à propos et évaluent les actions des autres personnes*, et que ces évaluations ont des conséquences pour les interactions futures (Boehm, 1999). [...] (b) *Que les jugements moraux sont donc définis comme étant : des évaluations (bonnes ou mauvaises) des actions ou du caractère d'une personne qui sont faits en fonction d'un ensemble de vertus considérées comme obligatoires par une culture ou une sous-culture* (Haidt 2001, p. 817-18, emphase ajoutée).

Dans cet extrait, Haidt maintient que la définition du jugement moral qu'il emploie se fonde sur le fait que (a) nous partageons avec les autres nos opinions à propos de l'action ou du caractère des individus. Ensuite, il définit le jugement moral comme étant (b) une évaluation des actions qui est faite sur la base d'un système de normes déterminant ce qui est conçu comme vertueux par les membres d'un groupe ou un sous-groupe culturel. Il me semble donc que pour Haidt, la moralité et la capacité de formuler de « vrais » jugements moraux requièrent plus que la simple capacité de ressentir des émotions morales. En d'autres mots, bien que pour Haidt la moralité trouve son origine dans les intuitions et les émotions

morales, peut-être que dans sa théorie les processus cognitifs plus réflexifs seraient nécessaires pour formuler un « vrai » jugement moral. Si notre intuition est juste, nous pourrions illustrer les premiers trois passages du schéma de Haidt de la façon suivante :

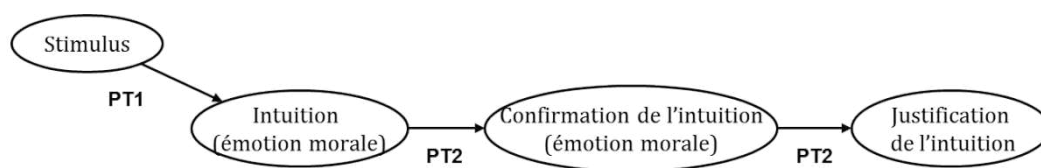


Fig. 1.3.3 PT1 et PT2 dans le modèle socio-intuitionniste de Haidt

Selon mon interprétation du modèle de Haidt, la réponse automatique et émotive au stimulus est endossée par les PT2 et ensuite défendue devant autrui à travers d'autres processus rationnels. Cette interprétation me semble aussi plus cohérente avec le modèle de processus duaux conçu par Kahneman et auquel Haidt fait explicitement référence. En effet, pour Kahneman, comme nous avons vu précédemment, le S2 est censé vérifier les réponses du S1 pour ou bien les confirmer ou bien les corriger. Il reste que Haidt n'est pas très clair à ce sujet.

Cela étant dit, nous pouvons maintenant examiner plus en détail l'aspect du modèle socio-intuitionniste qui nous intéresse : les processus de correction des jugements que nous pouvons mettre en place délibérément.

1.3.3 Réviser un jugement

Comme je l'ai mentionné dans l'introduction, dans ce mémoire j'essaie de répondre à la question suivante : serait-il possible de rendre les mécanismes de correction de nos jugements plus rapides et automatiques (question 1) ? Pour répondre à une telle question, un modèle de jugement moral doit être esquissé et pour ce faire, il est nécessaire de répondre à une autre question : quelles opérations mentales sont employées dans la révision des jugements moraux (question 2) ?

Comme nous avons vu, le modèle socio-intuitionniste de Haidt nous permet de répondre à cette question. En effet, pour Haidt nos jugements moraux peuvent être corrigés par deux types de processus : le raisonnement logique et le changement de perspective.

Dans la présente section, j'argumenterai afin de montrer qu'employer le changement de perspective afin de modifier nos jugements moraux est avantageux, puisque, selon le modèle de Haidt, les jugements auxquels le changement de perspective donne lieu nous motiveraient à l'action. Enfin, je montrerai que Haidt n'a pas exploré la possibilité de rendre automatique ce processus de correction des jugements.¹³ En d'autres mots, le modèle socio-

¹³ Au contraire, Haidt semble soutenir que nous pouvons changer nos réponses intuitives très facilement dans les années du développement où le cortex préfrontal (qui semble être la zone du cerveau la plus impliquée dans la formulation des jugements moraux) est très plastique, c'est-à-dire pendant l'adolescence. À l'âge adulte, Haidt semble exclure, sinon considérer très difficile, que nous puissions arriver à changer nos intuitions. Il affirme à ce propos que : « When an American adult later travels in Orissa, he may know how rules of purity and pollution govern the use of space, but he knows these things only in a shallow, factual, consciously accessible way; *he does not know these things in the cognitive/affective/motoric way that a properly enculturated Oriya knows them.* [...] People can acquire explicit propositional knowledge about right and wrong in adulthood, but it is primarily through participation in custom complexes (Shweder, et al., 1998) involving sensory, motor, and other forms of implicit knowledge (Fiske, 1999; Lieberman, 2000; Shore, 1996), that are shared with one's peers *during the sensitive period of late childhood and adolescence* (Harris, 1995; Huttenlocher, 1994; Minoura, 1992) that one comes to feel, physically and emotionally (Damasio, 1994; Lakoff & Johnson, 1999), the self-evident truth of

intuitionniste ne nous permet pas de répondre à la question qui nous intéresse au premier plan dans ce mémoire, à savoir, s'il est possible de rendre automatiques les processus qui nous permettent de réviser nos jugements moraux.

1.3.3.1 Le raisonnement logique et le changement de perspective

Comme j'ai mentionné précédemment, selon le modèle socio-intuitionniste, bien que la raison (PT2) soit souvent à la merci des intuitions (PT1), elle peut parfois prendre le contrôle. Plus précisément, selon Haidt, nous pouvons réviser nos jugements par deux types de PT2 : le raisonnement logique (flèche 6 dans la figure) et le changement de point de vue sur une situation (flèche 7).

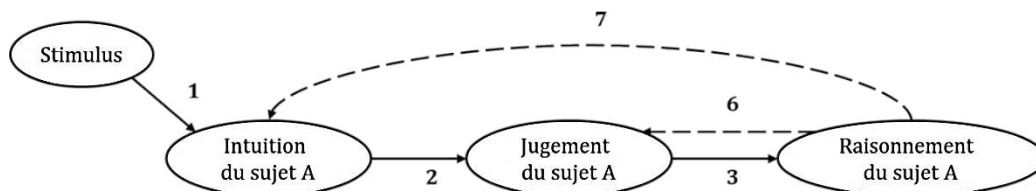


Fig. 1.3.3.1 Les processus de correction des jugements moraux dans le modèle socio-intuitionniste

Dans un premier cas, selon Haidt le PT2 qui nous permet de modifier notre jugement est le raisonnement logique (flèche 6). Voici comment il explique cette possibilité :

Les gens peuvent parfois raisonner jusqu'à un jugement par (a) *la simple force de la logique*, qui renverse [*overriding*] leur intuition initiale. Dans de tels cas, le

moral propositions » (Haidt 2001, p. 833-834, emphases ajoutées). Pour une perspective contraire au sujet de la plasticité émotionnelle et de la possibilité de changer nos intuitions voir : Faucher et Tappolet (2008).

raisonnement est réellement causal et ne peut être considéré comme étant « l'esclave des passions ». Cependant, des tels raisonnements sont posés comme étant rares se produisant principalement dans les cas où l'intuition initiale est faible et que la capacité de traitement [*processing*] est élevée. Dans des cas où le jugement raisonné entre en conflit avec un fort jugement intuitif, une personne aura une « attitude duelle » [*dual attitude*] (Wilson, Lindsey, & Schooler, 2000) dans laquelle (b) *le jugement raisonné est exprimé verbalement, même si le jugement intuitif existe toujours sous la surface* (Haidt 2001, p.819 emphase et énumération ajoutées).

Dans cet extrait, Haidt affirme que la « pure force de la logique » nous permet soit de rejeter le jugement qui suit de l'intuition, mais seulement lorsque celle-ci est faible, soit de formuler un autre jugement qui ne la neutralise pourtant pas lorsque l'intuition initiale est forte. En d'autres mots, dans le dernier cas, la réflexion qui révisé notre jugement ne change pas l'intuition initiale (chargée émotionnellement) et nous émettrons un nouveau jugement qui ne sera pas cohérent avec l'émotion qui le sous-tend.

À la lecture de ce passage, il est difficile de comprendre de quelle façon nous pouvons corriger un jugement par le raisonnement logique. En effet, le jugement ne semble pas pouvoir être illogique en soi. Ainsi, j'estime que l'élément qui pourrait être illogique n'est pas donc le jugement en tant que tel, mais sa justification. Nous pouvons penser à deux types de situations dans lesquelles le raisonnement logique peut modifier le jugement : (1) lorsque nous n'arrivons pas à trouver une justification à notre jugement (comme dans le cas de l'inceste que j'ai mentionné précédemment) ; (2) lorsque nous nous apercevons que la justification que nous élaborons pour le jugement comporte des inconsistances logiques. Étant donné que nous avons déjà expliqué ce qui se passe dans le premier cas (voir la section 1.3.2.1 et l'exemple de l'inceste), et que Haidt n'offre pas une explication plus détaillée du deuxième cas, je vais en donner un exemple. Supposons que nous avons un collègue, Sam, qui a une très mauvaise hygiène personnelle. On est donc dégoûté par Sam et ce sentiment négatif peut influencer nos jugements moraux envers lui. Supposons, par exemple, que nous apprenons que Sam ne paie pas ses impôts. On va dire que « Sam est une mauvaise

personne », mais ce jugement découle au final du dégoût suscité par la mauvaise hygiène de Sam. On cherchera ensuite une justification pour le jugement que nous avons formulé et nous pourrions aboutir à l'énoncer de la façon suivante : « Si quelqu'un paie ses impôts alors il est une bonne personne, mais Sam ne paie pas ses impôts, donc il n'est pas une bonne personne ». Or, ce raisonnement est un sophisme connu sous le nom de « négation de l'antécédent ». Imaginons que nous nous rendions compte de l'erreur logique que nous venons de commettre. Il serait possible de corriger ce jugement moral erroné et dire que nous n'avons pas assez d'information sur Sam pour le juger moralement. Pourtant, selon Haidt, l'intuition qui sous-tendait le premier jugement (le dégoût dans ce cas) ne changerait pas, il existerait toujours en parallèle de ce nouveau jugement.

Dans le deuxième cas, le raisonnement produit une nouvelle intuition en causant indirectement un nouveau jugement (flèche 7). Haidt affirme que :

La méthode la plus largement discutée pour générer de nouvelles émotions est (a) la prise de rôle [*role taking*] (Selman, 1971). En se mettant simplement dans les souliers d'une autre personne nous pouvons sentir instantanément de la douleur, de la sympathie ou d'autres réponses émotionnelles indirectes [*vicarious*] [...] Une personne en vient à voir une question ou un dilemme de plus d'un côté et (b) par le fait même, fait l'expérience de plusieurs intuitions en compétition. Le jugement final peut être déterminé [...] en suivant (c) l'intuition la plus forte (Haidt 2001, p. 819-20, énumération ajoutée).

Le raisonnement qui déclenche la nouvelle intuition consiste dans (a) le changement de perspective à partir de laquelle nous analysons une situation donnée. Chaque nouvelle intuition qui découle du raisonnement (b) entre en compétition avec les intuitions précédentes, et (c) celle qui est plus « forte » constituera le contenu du jugement moral. Haidt n'approfondit pas cette idée, mais les affirmations précédentes entraînent plusieurs questions. Par exemple, il faudrait expliquer ce qui rend une intuition plus forte qu'une autre, mais

l'exploration de ce point n'est pas nécessaire pour la poursuite de notre argument.¹⁴

Le modèle socio-intuitionniste permet donc que le jugement soit corrigé grâce à deux types de processus délibérés : l'un qui consiste à vérifier que les justifications de nos jugements respectent les règles logiques (flèche 6); l'autre qui implique un changement de point de vue sur la situation en question (flèche 7).

Conformément à la théorie de Haidt, l'usage de la logique serait donc une solution viable afin d'éviter que le biais affectif [*affective biais*] influence nos jugements moraux. Par contre, diverses recherches montrent que l'entraînement du raisonnement logique, l'apprentissage de la rhétorique et de la pensée critique ne sont pas des méthodes très efficaces pour diminuer l'influence des biais cognitifs sur nos jugements moraux dans le temps (Kenyon, T. et Beaulac, G. 2014). Des tels résultats sont à mon avis une raison suffisante pour diriger notre attention sur l'autre processus cognitif qui, selon Haidt, nous permet de corriger nos jugements moraux, soit le changement de perspective.

En outre, une des conséquences du modèle de Haidt, qu'il néglige cependant de mentionner, est que conformément à sa théorie, la correction des jugements à travers le changement de perspective influence le comportement (au moins dans le cas où l'intuition qu'elle produit est plus forte que l'intuition initiale),

¹⁴ La position intuitionniste de Haidt ne tient pas compte des cas semblables au suivant. Supposons qu'en discutant avec un ami j'affirme que « tuer est mal » sans ressentir *en même temps* l'intuition qui sous-tend généralement ce jugement. En effet, selon Haidt ou bien le jugement découle d'une intuition, ou bien il découle d'un raisonnement logique. Cela implique que, sauf dans le cas où nous révisons le jugement moral avec un raisonnement logique, dans toute autre situation notre jugement moral est sous-tendu par une intuition. Bref, le modèle de Haidt nécessite un lien synchronique entre l'intuition et le jugement. Toutefois, souvent nous prononçons des jugements moraux qui ne sont pas supportés par une intuition, bien qu'ils soient tout à fait cohérents avec des intuitions que nous ressentons habituellement. Le modèle de Haidt ne me semble pas pouvoir expliquer de telles situations.

contrairement à la correction à travers le raisonnement logique. Je vais consacrer la prochaine sous-section à l'explication de cet aspect du modèle de Haidt. Ce faisant, j'espère montrer que, dans le cas où le modèle de Haidt soit correct à ce sujet, il y aurait un avantage important à corriger nos jugements en adoptant un changement de point de vue.

1.3.3.2 Les intuitions motivent à l'action

Dans cette section, je montrerai pourquoi la théorie de Haidt prévoit qu'un jugement moral corrigé par un changement de point de vue sur la situation nous motive à l'action.

Comme j'ai mentionné précédemment, Haidt affirme que les intuitions morales sont constituées par des émotions, à propos de ces dernières Haidt affirme que :

Les émotions sont souvent analysées en composantes, tel qu'un événement éliciteur [*eliciting event*], une expression faciale, un changement physiologique, une expérience phénoménologique, et une motivation ou une tendance à l'action [...] Deux de ces composantes sont utiles pour identifier les émotions morales, puisqu'elles sont facilement liées aux intérêts de la société ou à celui d'autres personnes : les éliciteurs et les tendances à l'action (Haidt 2003, p. 853, emphase ajoutée).

Dans cet extrait, Haidt affirme que les émotions présentent plusieurs facettes, notamment, elles sont caractérisées par une motivation ou une tendance à l'action.¹⁵ De la même façon, dans la littérature philosophique sur les émotions, la

¹⁵ Une question demeure ouverte à propos des caractéristiques propres aux émotions : laquelle d'entre elles constitue un trait essentiel des émotions ? (Tappolet 2010, p. 326). Pour Haidt chaque émotion a une charge motivationnelle. Cependant, la question reste ouverte quant à savoir si la

motivation à l'action est généralement conçue comme étant ou bien un désir ou bien une disposition à un comportement (Tappolet 2010, p. 326).

Comme j'ai mentionné, les jugements corrigés à travers le raisonnement logique sont incompatibles aux intuitions qui le sous-tendent. Donc, ces jugements n'ont pas de charge émotive, ce qui fait que la motivation demeure du côté de l'intuition initiale qui n'a pas été changée.

Reprenons l'exemple que j'ai donné précédemment : nous avons changé notre jugement « Sam est une mauvaise personne » après avoir compris que sa justification se fondait sur un raisonnement fallacieux (la négation de l'antécédent). Supposons que ce jugement découle du fait que Sam nous dégoûte. Cette affection négative que nous éprouvons envers Sam qui constitue le contenu de l'intuition. Cette intuition nous incite à juger Sam plus sévèrement. Selon le modèle de Haidt, modifier notre jugement à travers le raisonnement logique n'effacera pas l'affection qu'on ressent envers Sam. Conséquemment, même si nous pourrions dire publiquement « Je ne connais pas assez Sam pour en dire quoi que ce soit », nous continuerons à nous méfier de lui.

Au contraire, selon le modèle de Haidt, les jugements corrigés par un changement de perspective sur la situation découlent d'une nouvelle intuition. Ainsi, changer le jugement de cette manière pourrait éventuellement susciter une intuition différente de la première qui nous motivera à agir différemment.

En conclusion, le modèle de Haidt prévoit que corriger nos jugements sans changer l'émotion qui les sous-tend n'aurait pas d'effet sur le déroulement de nos vies sociales, puisqu'ils ne motiveront pas à agir différemment (ou pas durablement, l'intuition restant toujours susceptible de venir interférer avec nos jugements raisonnés).

charge motivationnelle (qu'elle soit un désir ou une disposition au comportement) constitue un caractère essentiel des émotions.

1.4 Conclusion

Est-il possible d'apprendre à corriger automatiquement nos jugements moraux afin d'éviter qu'ils soient influencés par le biais affectif ? Comme je l'ai mentionné dans l'introduction à ce chapitre, donner une réponse plausible à cette question requiert que nous répondions préalablement à une autre question. Il faudrait tout d'abord déterminer quels processus cognitifs nous permettent de modifier nos jugements moraux. Pour ce faire, dans ce chapitre j'ai présenté deux modèles de jugements moraux qui sont très influents dans la littérature en psychologie sociale, celui de Greene et celui de Haidt.

Greene nous encourage simplement à douter de nos jugements et à y réfléchir davantage pour faciliter la formulation de jugements moraux conséquentialistes. Ainsi son modèle propose que la réflexion ne joue pour tout rôle que celui de permettre la révision de nos jugements moraux (pas la modification de nos intuitions). Comme j'ai essayé de le montrer, bien que les études amenées par Greene nous donnent une bonne base empirique pour construire un modèle du jugement moral, sa suggestion présente plusieurs défauts. Tout d'abord, pour l'adopter nous devrions au moins savoir que nos jugements moraux sont influencés par des biais cognitifs. De plus, son modèle du jugement moral n'explique pas pourquoi nous serions motivés à corriger nos jugements moraux. En outre, sa méthode ne peut être appliquée que dans les cas où l'action qui est jugée cause un mal physique à quelqu'un. Enfin, Greene semble exclure que les processus de révision des jugements moraux puissent devenir automatiques par l'habitude ou par un processus semblable. En effet, il affirme que nous ne devrions pas faire confiance aux processus automatiques et qu'il faudrait employer plutôt les processus réflexifs (Greene 2014, p. 34).

Le modèle de Haidt, au contraire, est applicable à beaucoup plus de situations morales. Selon Haidt, nous pouvons réviser nos jugements à travers le raisonnement logique ou par un changement de perspective. Haidt n'explore cependant pas la possibilité de rendre automatique des tels mécanismes de correction des jugements. Cela peut paraître d'autant plus curieux lorsqu'on se rappelle que son modèle (comme celui de Greene) a été inspiré par la théorie des processus duaux élaborée par Kahneman, théorie selon laquelle les PT2 peuvent devenir des PT1 avec le temps et l'habitude. Or, comme j'ai mentionné, les méthodes de *debiasing* par l'entraînement du raisonnement logique ne semblent pas donner des résultats satisfaisants. On pourrait penser qu'il en sera de même dans le domaine moral. Cependant, selon Haidt, le raisonnement logique n'est qu'un des deux processus qui interviennent pour corriger nos jugements moraux. Nous pourrions donc explorer la possibilité d'améliorer notre capacité à changer de point de vue, afin de diminuer l'influence que le biais affectif a sur nos jugements moraux.

Dans le prochain chapitre, j'essaierai de montrer que, selon l'alternative que je propose et qui est inspirée par la théorie morale de David Hume, nous pouvons entraîner un certain type de *sympathie* grâce à laquelle nous pouvons changer nos jugements moraux ou prévenir les erreurs de jugement.

CHAPITRE II

UNE SOLUTION HUMIENNE

Au tout début de son *Enquête sur les Principes de la Morale* (EPM), Hume adopte une position intermédiaire dans le débat philosophique où s'opposent ceux pour qui nos jugements moraux dérivent de la raison et ceux qui croient qu'ils se fondent sur un sentiment (EPM 1.3).¹⁶ À ce propos, Hume affirme que chacune de ces deux positions philosophiques pourrait être solide et satisfaisante et qu'autant la raison que les sentiments nous permettent de distinguer les vices et les vertus. Plus précisément, Hume affirme qu'au final c'est le sentiment qui détermine les jugements moraux, tout en nuancant de la façon suivante que :

[...] afin de préparer le terrain pour un [a] tel sentiment, et de lui permettre un juste discernement de son objet, on constate qu'il doit souvent être nécessairement précédé de force raisonnements, qu'il faut effectuer des distinctions fines et tirer des conclusions justes, établir des comparaisons entre faits éloignés, examiner des relations compliquées, déterminer et assurer des faits d'ordre général. [...] dans les beaux-arts, *il est nécessaire de faire largement appel au raisonnement, si l'on veut éprouver* [b] *les sentiments appropriés*, et une fausse appréciation peut le plus souvent être corrigée par l'analyse et la réflexion. Nous avons de bonnes raisons de conclure que la beauté morale participe largement de cette dernière catégorie et exige l'assistance de nos facultés intellectuelles, afin d'en tirer une influence convenable sur l'esprit humain (EPM 1.9, emphases ajoutées).

Dans ce passage, Hume affirme, d'une part, que [a] les jugements moraux dépendent d'un sentiment qui découle de certains raisonnements, et d'autre part

¹⁶ Dans ce chapitre, le terme « jugement moral » continuera à indiquer l'évaluation d'un individu, de ses actions, de ses émotions ou de ses attitudes comme étant *bonnes* ou *mauvaises*, *vertueuses* ou *vicieuses*. Il faut pourtant préciser que dans le langage humien le terme « jugement moral » désigne un type de jugement évaluatif qui se fonde sur un *sentiment moral* et qui a une prétention d'universalité, comme on verra plus en détail par la suite. Afin d'établir une continuité entre les chapitres de ce mémoire et pour pouvoir comparer la théorie humienne de la correction des jugements avec les théories analysées précédemment, il est nécessaire que le terme « jugement moral » désigne le même type d'évaluation.

que [b] le *sentiment approprié* découle seulement de *raisonnements fins, justes et assurés*. Cela implique que nous pouvons formuler deux types de jugements moraux : (1) des jugements moraux fallacieux qui découlent de sentiments inappropriés ; (2) des jugements moraux justes qui trouvent leur origine dans des sentiments appropriés.¹⁷ En effet, suite au passage que je viens de citer, nous savons que pour Hume certains sentiments envers l'objet du jugement moral sont appropriés alors que d'autres ne le sont pas.

Dans la suite de ce chapitre, j'exposerai d'abord ce qui peut être l'objet du jugement moral et comment les sentiments (appropriés ou non) qui constituent le contenu de nos jugements sont produits par la sympathie. Ensuite, j'illustrerai quels sont les erreurs qui nous mènent à ressentir les sentiments inappropriés et enfin comment, selon Hume, nous pouvons réparer ces erreurs afin de ressentir les *sentiments appropriés*.

¹⁷ Or, on pourrait soulever l'objection suivante à une telle lecture du texte : Hume dit très clairement que le contenu de nos évaluations morales réside dans le sentiment que notre inférence déclenche. Une inférence peut comporter une mésinterprétation des actions des autres sur le plan de leurs motivations et être conséquemment erronée. Au contraire, l'évaluation morale qui se fonde sur cette inférence ne sera pas fallacieuse, étant donné que, pour Hume, nous ferions une erreur de catégorie en affirmant qu'un sentiment est vrai ou faux. Il s'en suivrait donc qu'une évaluation morale n'est jamais vraie ou fausse et qu'elle n'est même pas erronée. Cependant, comme nous l'avons vu, bien que pour Hume le sentiment qui détermine le jugement moral ne puisse pas être erroné, il peut être inapproprié. Selon l'auteur, certains sentiments constituant le contenu de nos évaluations morales sont plus appropriés que d'autres, et pour ressentir le sentiment approprié (à savoir le *sentiment moral*), il faut aussi saisir adéquatement les causes des actions (j'expliquerai davantage ces dernières affirmations dans la troisième section du chapitre). À ce sujet, pour faire suite à la note 16, si on voulait être fidèle au langage humien, ceux que j'appelle « jugements moraux justes » seront les jugements moraux et ceux que j'appelle « jugements moraux fallacieux » seront plutôt des jugements évaluatifs qui se fondent sur un sentiment inapproprié et donc pas du tout des jugements moraux. Sur la question du réalisme moral humien, voir : Pitson (1989). Pour une interprétation selon laquelle Hume soutient une forme de réalisme moral, voir : Remow (2009) aussi bien que Norton (1984). Pour l'idée selon laquelle les sentiments peuvent être appropriés ou pas, voir : D'Arms et Jacobsen (2000).

2.1 L'objet du jugement moral

Dans le cadre de la théorie morale de Hume, les évaluations morales portent sur les motivations des agents. À ce propos Hume affirme :

Il est évident que, lorsque nous louons des actions, nous ne considérons que les motifs qui les engendrent [...] Il apparaît, par conséquent, que *toutes les actions vertueuses tirent uniquement leur mérite de motifs vertueux et sont considérées purement comme des signes de ces motifs* (T 3.2.1.2 et 4).

Donc, afin de juger moralement un individu il faut tout d'abord saisir ce qui motive son action. Hume consacre une section entière du *Traité* à montrer que seules les passions peuvent motiver notre volonté à agir (T 2.3.3).¹⁸ Plus précisément, il soutient que seuls les désirs ou les aversions (qui sont pour lui des passions) poussent la volonté à provoquer une action du corps ou de l'esprit (T 2.3.9.7).¹⁹ Ainsi, si pour juger moralement nous devons tout d'abord saisir ce qui a motivé l'action, et si la motivation consiste toujours en un désir ou en une aversion, alors la première étape de l'évaluation morale consiste à saisir quel désir ou quelle aversion a motivé l'action.

Il faut aussi noter que pour Hume nous ne jugeons pas seulement des passions qui contiennent un désir ou une aversion et qui se manifestent à nos yeux à travers les actions. Nous pouvons aussi juger des passions qui se révèlent dans les expressions, les attitudes et la conversation. Comme les actions, ces signes

¹⁸ Il vaut la peine de souligner que pour Hume la volonté est « l'impression interne que nous sentons et dont nous sommes conscients, lorsque nous suscitons sciemment un nouveau mouvement de notre corps ou une nouvelle perception de notre esprit » (T 2.3.1.2). D'ailleurs, la volonté, dit-il, « intervient », seulement nous considérons que l'objet de notre désir (ou de notre aversion) peut être atteint (ou évité) « par une action quelconque de l'esprit ou du corps » (T 2.3.9.7).

¹⁹ Je ne pourrais pas rendre compte ici de façon exhaustive des détails concernant la théorie des passions présentée par Hume. Il suffira de dire que nous pouvons avoir deux types de désirs envers les autres : le désir de leur bonheur ou le désir de leur malheur. La bienveillance et la rage sont les passions qui contiennent ces désirs. Elles accompagnent l'amour et la pitié (T 2.2.6.3 et T 2.2.7) et la haine et la malice respectivement (T 2.2.8.12).

extérieurs peuvent guider l'esprit à conclure que quelqu'un est affecté par une passion, qui peut se révéler louable ou blâmable (T 3.3.2.10). Afin d'alléger le texte, je parlerai dorénavant seulement des actions en tant que signes des passions animant un individu, en incluant dans ce terme les autres types de signes extérieurs des passions que je viens de mentionner.

En raison de ce qui précède, je peux affirmer que pour Hume les passions sont la *cause* des actions.²⁰ Or, selon Hume, l'opération mentale qui nous permet de déduire une cause à la vue de son effet (et vice-versa) est l'*inférence* (T 1.3.6.1).²¹ Conséquemment, nous « lisons » les intentions des autres en inférant la passion à la vue de l'action que nous supposons être son symptôme.

De plus, la psychologie humienne veut que l'esprit puisse raisonner par inférences seulement lorsque deux idées sont associées causalement dans l'imagination. Afin de bien expliquer cette première étape de l'évaluation morale, j'esquisserai une explication de comment, selon Hume, deux idées deviennent associées dans l'esprit selon la relation de causalité.²² Cela nous permettra de rendre compte de comment, selon Hume, nous arrivons à ressentir le sentiment qui constitue le contenu du jugement moral.

²⁰ Or, il est important de souligner que pour Hume une action n'est pas possible si le désir n'est pas accompagné par certaines opérations de la raison. Notamment, il faut qu'on pense que l'objet désiré existe et il faut qu'on raisonne sur les moyens qui nous assureraient d'atteindre notre objectif (T 3.1.1.12).

²¹ David Hume utilise le terme anglais « *inference* » qui signifie : « Mener à (quelque chose) comme une conclusion ; impliquer » (« infer, v. » *OED Online*. Oxford : Oxford University Press, March 2014. Web. 22 March 2014. Def.4).

²² Pour Hume, deux idées sont en relation à la fois lorsqu'elles sont associées (et elles le deviennent avec la coutume comme nous le verrons sous peu) et lorsqu'elles ne sont pas associées, mais simplement mises en comparaison par l'imagination (T 1.1.5.1). Par exemple, l'imagination peut arbitrairement comparer l'idée d'un cheval et l'idée d'une licorne selon leur ressemblance. Ainsi, elle chercherait les points communs et les différences entre ces deux idées. Elle peut faire cela même lorsque l'idée du cheval et l'idée d'une licorne ne sont pas associées dans l'imagination. Ce n'est que si nous recevons constamment et habituellement les impressions de l'un conjointe à l'impression de l'autre, que l'idée de l'une deviendra « unie par les relations » (T 2.1.4.2) à l'idée de l'autre, comme nous le verrons sous peu. Pour plus d'informations quant aux différentes manières que l'imagination a pour mettre en relation les idées voir : T 1.1.5 et T 1.3.1.1.

2.1.1 L'association d'idées

Hume affirme que nos inférences permettent de tirer l'effet à partir de la vue de sa cause (et vice-versa). Par exemple, si je vois de la fumée au loin, je conclus qu'elle est causée par le feu, même si je ne peux pas voir de feu. Il affirme aussi que cette opération mentale dépend de l'association causale de l'idée de la fumée et l'idée du feu dans l'esprit. Au cours de cette section, je vais montrer que pour Hume chaque évaluation morale trouve son origine dans cette *association causale* entre l'idée d'une passion et l'idée d'une action, dont la première est considérée comme étant la cause de la deuxième. Pour cette raison, il est tout d'abord nécessaire de mentionner que pour Hume c'est l'expérience qui provoque l'association de l'idée d'une action à l'idée une passion dans notre esprit.²³ Cela étant dit, il est maintenant opportun de faire une courte parenthèse afin de clarifier le sens que les termes « idée » et « passion » ont dans la psychologie humienne.

Premièrement, d'après la théorie humienne, les perceptions se divisent en impressions et idées (Figure 2.2.1) (T 1.1.1).²⁴

²³ À ce propos, selon Hume, il n'y a rien dans les causes qui nous permette de prévoir avec le seul raisonnement quels effets en suivront (EEH 4.7). En outre, il affirme que contrairement aux idées des sons, des saveurs ou des couleurs, dont l'impression est perçue par les sens et d'où ces idées découlent, il est impossible pour nous de retrouver la perception sensorielle qui génère en nous l'idée de causalité. Hume finira par montrer que c'est le sentiment qui accompagne nos croyances qui constitue l'impression de réflexion sur laquelle se fonde notre idée de la causalité. L'impression qui génère l'idée de la causalité est donc un produit de notre esprit (T 1.3.2) et elle consiste dans le sentiment qui accompagne toujours nos croyances (EEH 7.28).

²⁴ Selon David Fate et Mary J. Norton, pour Hume chaque élément mental qui est l'objet de notre attention est une perception de l'esprit. Seulement en ce sens pouvons-nous dire que l'esprit « perçoit » les événements (p. 426).

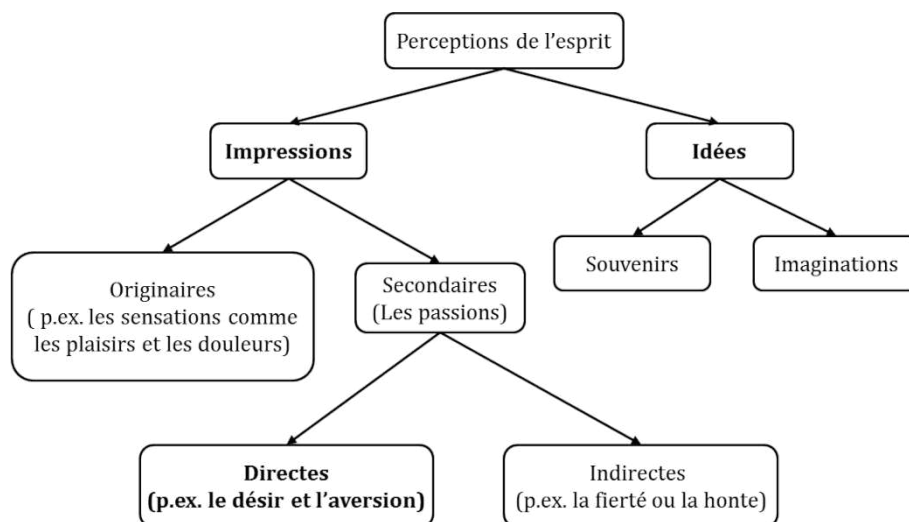


Fig. 2.2.1 Les types de perceptions dans la théorie psychologique de David Hume

25

Les impressions se distinguent entre impressions originaires et impressions secondaires.

Les impressions originaires sont les *sensations* suscitées de l'intérieur (du mouvement des « esprits animaux », c'est-à-dire par des processus physiologiques) ou de l'extérieur du corps (des objets) : le plaisir et le malaise en sont deux exemples (T 1.1.2.1 /2.1.1). Ainsi, si je me brûle en touchant le feu, cette douleur est une impression originaires qui trouve son origine à l'extérieur de mon corps, tandis que si j'éprouve des maux de ventre, cette douleur est une impression originaires qui provient de l'intérieur de mon corps.

Quant aux impressions que Hume appelle *secondaires*, elles sont causées par les sensations, le plus souvent par l'intermédiaire d'une idée : l'aversion pour le feu qui trouve son origine dans l'idée de la douleur (causée à son tour par la brûlure) est une impression secondaire, c'est-à-dire une passion (T 2.1.1.1-2).

²⁵ Pour l'instant je n'ai pas inséré les sentiments moraux et le sentiment esthétique dans ce diagramme puisqu'on verra par la suite et plus en détail comment ils s'y inscrivent.

Enfin, Hume distingue les passions directes et les passions indirectes en affirmant que les premières surviennent immédiatement des sensations de plaisir ou de malaise. Les désirs et les aversions en sont un exemple : nous pouvons désirer manger un morceau de chocolat puisqu'on a expérimenté que cela nous plaît. Tandis que les passions indirectes découlent de ces sensations de douleur ou de plaisir, jointes à d'autres qualités de leurs objets.²⁶

Deuxièmement, Hume distingue les impressions des idées en fonction de leur force. Ainsi, alors que les impressions sont des perceptions très fortes et vivaces, les idées en sont des copies affaiblies qui dérivent des impressions (T 1.1.1.3/1.1.1.8-12).²⁷ Par exemple, tant que je ressens de la peur, cette perception est une impression. Lorsque je cesse d'être épouvantée, le souvenir de la peur demeure dans mon esprit sous la forme d'une idée. Les souvenirs sont les idées qui demeurent dans notre esprit juste après en avoir reçu l'impression.²⁸ Puisqu'ils

²⁶ Je ne pourrais pas parler davantage, dans ce mémoire, des différents types de passions dont Hume fait mention. Pour plus d'information à ce sujet voir : T 2.1-2-3 ou *A Dissertation on the Passions and The Natural History of Religion*, Oxford : Clarendon Press, 2007.

²⁷ Or, par souci de précision, Hume distingue aussi les perceptions simples et complexes. Les perceptions simples ne peuvent pas être divisées, contrairement aux perceptions complexes. Ainsi, il y a des idées et des impressions simples aussi bien que des idées et des impressions complexes. Notamment, l'impression d'une couleur est une perception simple, tandis que l'impression d'une ville est une perception complexe. En effet, nous ne pouvons pas diviser l'impression de la couleur en des impressions plus simples, tandis qu'il est aisé de diviser l'impression d'une ville en impressions plus simples, telles que les impressions de ses couleurs, de ses formes, de ses odeurs, etc... Hume spécifie à ce propos que toutes les idées simples sont des copies affaiblies des impressions simples (sauf dans un cas très rare qu'il mentionne au T 1.1.1.11, mais que je ne résumerai pas ici). Au contraire, les idées complexes ressemblent aux impressions complexes, mais il n'est pas nécessaire qu'elles en soient des copies. Cela permet aussi d'expliquer pourquoi nous pouvons avoir des idées des villes que nous n'avons jamais visitées (T 1.1.1.3-5). Voir aussi Schmicking 2004 et Hausman 1988 sur cette distinction.

²⁸ Les souvenirs sont aussi ordonnés conformément aux impressions qui les ont suscités (p. ex. je me rappelle que le chat a miaulé juste après que j'ai marché sur sa queue). Pourtant, selon Hume, ce qui nous permet de distinguer une idée de la mémoire d'une idée de l'imagination n'est pas leur ordre, « étant dans l'impossibilité de rappeler les impressions passées afin de les comparer avec nos idées présentes et voir si leur arrangement est exactement semblable » (T 1.3.5.3). En dépit de ceci, nous pouvons distinguer une idée de la mémoire grâce au sentiment [*feeling*] (T 1.3.5.4) qui la caractérise. À ce propos, Hume affirme que : 1) ce sentiment dépend du niveau de force et de

découlent immédiatement des impressions, les souvenirs sont des idées très vivaces et fortes. En revanche, plus le temps passe, plus cette empreinte de la sensation qui est le souvenir, perd de sa vivacité et de sa force et se change en imagination (T 1.3.5.5). Il faut ici souligner que chaque idée perd graduellement sa vivacité. Cependant, l'écart entre les idées et les impressions n'est pas infranchissable. Il est toujours possible qu'une idée acquière assez de force pour qu'elle devienne une impression à nouveau, notamment, grâce à l'association d'idées.²⁹ Nous pouvons donc maintenant revenir à l'explication de la façon dont l'idée d'une action devient associée dans l'esprit à l'idée d'une passion, en se souvenant que lorsque ces deux idées sont associées dans l'esprit, cette association nous permet d'inférer quelle passion a causé une action.

Selon Hume, l'idée d'une passion et l'idée d'une action deviennent associées dans l'esprit selon la relation de causalité.³⁰ Nous pouvons comprendre comment l'esprit associe selon la causalité ces deux idées à l'aide de l'analogie selon laquelle l'esprit est comme une toile sur laquelle les expériences laissent leurs traces.

Deux événements, ma passion *P* et mon action *A* dessinent deux images très vivaces (impressions) sur la toile. Une fois que les événements cessent, les images (idée de *P* et idée de *A*) restent comme souvenirs dans l'esprit. Mais avec le temps

vivacité de l'idée qui le suscite ; et 2) qu'il nous pousse à donner notre assentiment à la perception qu'il accompagne.

²⁹ Hume ne donne pas des informations supplémentaires qui nous aident à déterminer quel niveau de force et de vivacité permet aux idées de devenir des impressions. En outre, dans l'*EEH*, il est beaucoup plus prudent lorsqu'il affirme que les idées peuvent atteindre un degré de force et de vivacité tel qu'on pourrait presque [*almost*] dire qu'on en a une impression (*EEH* 2.1). Il semble en effet affirmer que c'est seulement dans un état de folie ou de maladie que l'esprit peut nous présenter une idée ayant le même degré de force et de vivacité d'une impression. Voir aussi Landy (2006) sur cet aspect de la théorie humienne.

³⁰ La causalité est un des trois types de qualités associatives inhérentes aux perceptions qui mènent l'esprit à associer les idées. Les autres deux sont la contiguïté et la ressemblance, auxquelles je vais faire référence dans la deuxième section de ce texte.

elles perdent graduellement leurs couleurs. Or, la répétition d'expériences semblables ou analogues (T 1.3.12.25) donne à nouveau des couleurs aux images laissées par les expériences précédentes ; le dessin acquiert ainsi plus de force et de vivacité.³¹ En même temps, *l'habitude* de la cooccurrence d'événements semblables ou analogues à *P* et *A* forme une ligne unissant leurs traces dans l'imagination, qui symbolise ici leur association. Lorsque les deux idées sont associées dans l'imagination, notre esprit en perçoit la *connexion nécessaire* (T 1.3.6.14).³²

L'effet principal de l'association de deux idées *P* et *A* dans l'esprit est que l'impression de l'une suscite l'apparition de l'idée de l'autre dans l'esprit. Comme je l'ai mentionné précédemment, Hume appelle *inférence* cette opération mentale qui permet à l'imagination de passer d'une impression à l'idée qui lui est associée (T 1.3.6.15).³³ Par souci de précision, je veux ici souligner que l'inférence est une

³¹ Le lecteur pourra remarquer que je fais ici référence au premier livre du *Traité*, dans lequel Hume ne parle pas encore des inférences concernant la relation entre les actions et les intentions des agents. Il est pourtant possible d'utiliser ces passages afin d'expliquer comment l'idée d'une passion et l'idée d'une action deviennent associées dans l'esprit et comment nous pouvons inférer la passion à la vue de l'action, puisque Hume affirme clairement que : « [...] pour juger les actions des hommes, il nous faut procéder selon les mêmes maximes que dans nos raisonnements qui portent sur les objets extérieurs » (T 2.3.1.12), dont Hume parle dans les sections auxquelles je fais référence ici.

³² Il faut ici souligner que pour Hume la relation de *connexion nécessaire* entre les deux événements est une perception qui ne dérive pas d'une seule impression provenant du monde extérieur. En d'autres mots, selon Hume l'événement *x* renvoie à l'événement *y* seulement parce qu'on a fait expérience de leur lien constant. La connexion nécessaire est un produit de l'habitude et dépend de l'association des idées dans l'imagination. Dans ce contexte, « nécessaire » ne désigne donc pas une relation métaphysique propre à des entités existant indépendamment de l'esprit. Dans les mots de Hume : « Si des objets n'étaient pas entre eux en conjonction régulière et uniforme, nous ne parviendrions jamais à une idée de cause et d'effet » (T 2.3.1.4).

³³ Il faut ici noter que Hume n'est pas clair au sujet des tâches qui sont propres à l'imagination, à la mémoire, à l'intellect [*understanding*] et à la raison [*reason*]. En outre, il ne définit pas clairement et avec précision la nature de ces facultés. Nous pouvons ici nous baser sur une note en bas de page que l'auteur ajoute dans le *Traité*, où il affirme que : 1) l'imagination et la mémoire se distinguent sur la base de la vivacité et de la force de leurs idées (comme nous verrons sous peu); 2) la raison (qu'il appelle parfois intellect, p.ex. EEH 5.22) gère les raisonnements démonstratifs et probabilistes. En d'autres mots, la mémoire est la faculté qui recueille les idées plus fortes et vivaces, l'imagination et la raison /intellect gèrent les idées plus faibles, sauf que seule la raison ou l'intellect construisent des raisonnements démonstratifs ou probabilistes. Hume ajoute aussi que, lorsque la distinction entre la raison et l'imagination n'est pas mentionnée explicitement dans le

opération de l'esprit qui permet à une idée d'acquérir de la force et de la vivacité, et qu'elle est rendue possible par l'association de deux idées dans l'esprit. En effet, deux idées deviennent associées selon la causalité grâce à l'habitude que nous faisons de leur conjonction constante, et cette association fait en sorte que l'imagination puisse inférer une idée à partir de l'impression de l'idée qui lui est associée. Dans le cas particulier où nous « lisons » les intentions des autres agents, cette inférence implique que, si l'idée de *P* et l'idée de *A* sont associées dans l'imagination, une fois que l'esprit perçoit l'impression de *A*, cette impression transfère une partie de ses couleurs (force et vivacité) à l'idée de *P*. Ainsi, l'idée *P* se présente à l'esprit plus colorée (forte et vivace) qu'avant au point que nous aboutissons à croire que *P* a causé *A*. En effet, Hume affirme clairement qu'une *croyance* est le fruit d'une inférence fondée sur l'association d'idées par causalité (T 1.3.8.6).³⁴

Avant de passer à l'explication de certains autres aspects de la théorie morale de Hume, j'aimerais ajouter une remarque importante. Nous avons vu plus haut que

texte et lorsque le contexte ne permet pas de comprendre s'il faut entendre l'imagination dans un sens large ou plus étroit, il faut interpréter ses propos comme si l'imagination incluait la raison (donc le terme imagination indiquerait seulement la faculté qui gère les idées faibles) (T 1.3.9 note 22). Dans ce cas, il semble donc que le terme imagination indique cette faculté dans le sens plus large incluant la raison, étant donné que pour Hume l'imagination seulement en tant que raison/intellect met en place les raisonnements causaux (EEH 7.27).

³⁴ Pour l'auteur la croyance est une idée accompagnée d'un sentiment, tel qu'on donne notre assentiment à son contenu (T 1.3.7.7), elle nous persuade (T 1.3.9.6). Nous pourrions maintenant nous demander pourquoi je n'ai fait aucune mention des croyances dans la section précédente où j'ai distingué entre les différents types de perceptions de l'esprit. Il faut donc ajouter une précision afin d'éviter certaines confusions : Hume n'affirme pas que les croyances sont des types d'idées, mais plutôt qu'une croyance est une idée conçue d'une façon [*manner*] particulière (T 1.3.7.7) et qui est accompagnée par un sentiment. En suivant l'interprétation de Gorman M. (1993) je crois que pour Hume ces deux facettes de la croyance coïncident. Ainsi, le sentiment qui accompagne la croyance consiste en la façon dont nous la concevons : plus forte et vivace des autres idées de l'imagination. En outre, Hume affirme clairement que ce sentiment qui accompagne la croyance consiste en sa force et vivacité (T 1.3.7.7). Il affirme aussi que: « Belief is the true and proper name of this feeling; and no one is ever at a loss to know the meaning of that term; because every man is every moment conscious of the sentiment represented by it » (EEH 5.12).

l'association entre deux idées dans l'esprit est le produit d'une habitude à la cooccurrence constante de deux ou plusieurs événements. Sur ce point, il me semble que pour Hume, chaque fois qu'un événement qu'on croyait devoir suivre un autre ne se produit pas, l'habitude de leur cooccurrence s'affaiblit, de même que la *croyance* qui découle de l'association entre les deux idées (EEH 6.4). Hume affirme à ce propos que :

[...] quand nous n'avons pas observé un nombre de cas suffisant pour produire *une forte habitude*, ou quand ces cas sont contraires les uns aux autres, ou quand la ressemblance n'est pas exacte, ou quand l'impression présente est faible et obscure, ou quand l'expérience est dans une certaine mesure effacée de la mémoire, ou quand la connexion dépend d'une longue chaîne d'objets, ou quand l'inférence dérive de règles générales et n'y est pourtant pas conforme, dans tous ces cas, *l'évidence diminue* par la diminution de force et de vivacité de l'idée. (T 1.3.13.19, emphase ajoutée).

Ainsi peut-on dire que nous avons des *croyances* plus ou moins fortes concernant la passion qui aurait causé une certaine action, dépendamment du niveau de certitude qui accompagne nos croyances et qui, lui, dépend de l'habitude que nous avons eue de constater leur cooccurrence dans le passé.

Nous comprenons maintenant mieux en quoi, selon la théorie humienne, les jugements moraux dépendent de l'association entre l'idée d'une passion et l'idée d'une action, qui permet d'inférer les intentions des agents. Mais, comme nous avons précédemment vu, selon Hume, les jugements moraux reposent au final sur des sentiments. Pourtant, à cette étape de notre analyse, nous n'avons pas encore les éléments nécessaires pour saisir comment ces sentiments surgissent une fois que nous avons inféré quelle intention a motivé l'action d'autrui. Pour clarifier ce point, il faut introduire la notion de sympathie.

2.1.2 Le mécanisme de la sympathie

L'auteur présente ainsi le concept de *sympathie* dans le *Traité* :³⁵

Nulle qualité n'est plus remarquable dans la nature humaine, à la fois en elle-même et dans ses conséquences, que notre propension à sympathiser avec les autres et à recevoir par communication leurs inclinations et leurs sentiments, fussent-ils différents des nôtres, voire contraires aux nôtres (T 2.1.11.2, emphase ajoutée).

Dans ce passage, Hume définit le mécanisme sympathique comme étant un principe de communication des passions. Pour Hume, ce mécanisme psychologique qui nous permet d'attraper [*to catch*] (EPM 7.2) immédiatement les passions d'autrui est naturel, dans le sens qu'il agit avec constance dans notre esprit.³⁶ Selon Hume, il n'y a en effet qu'un seul cas où le mécanisme

³⁵ Dans le *Traité* aussi bien que dans l'*EPM*, Hume utilise le terme sympathie pour désigner : a) le sentiment qui résulte de la communication des passions, autrement dit, le sentiment qui est communiqué ; b) le mécanisme qui permet la communication des passions. Dans ce mémoire, je porte mon attention sur la sympathie comprise dans le second sens. À ce propos, il faut dire que pour Hume, le mécanisme sympathique fonctionne de plusieurs façons différentes. Je me bornerai à en décrire deux dans ce texte : en premier lieu dans cette section, j'expliquerai le fonctionnement de la sympathie limitée. Ensuite, dans la troisième partie de ce chapitre, je décrirai le fonctionnement de la sympathie étendue lorsqu'elle nous permet, selon Hume, de ressentir des sentiments appropriés.

³⁶ À ce sujet, je tiens à souligner que Hume tient compte des différents sens que le terme 'nature' peut avoir (T 3.1.2.7-11), ici je ne fais que présenter celui qui se prête le mieux pour comprendre pourquoi, pour l'auteur, la sympathie comme contagion émotionnelle est *naturelle*. En traitant des passions de l'orgueil et de l'humilité, Hume affirme qu'elles ont comme propriété naturelle et originelle celle de faire surgir dans notre esprit l'idée de *moi*, et il justifie ainsi ces affirmations :

Nul ne peut douter que cette propriété est [a] *naturelle*, compte tenu de la constance et de la stabilité de ses opérations [...] Que ce fait provienne [b] d'une qualité *originelle*, ou d'une impulsion primitive, nous apparaîtra tout aussi évident pour peu que nous considérons qu'il est la caractéristique distinctive de ces passions (T 2.1.3.2-3).

Conséquemment, pour Hume, d'une part, nous pouvons affirmer qu'une propriété est naturelle [a] si nous observons qu'elle opère en nous de façon stable et constante. D'autre part, une propriété est originelle [b] si elle fait partie de la constitution primaire de la nature humaine. Étant donné que Hume affirme que la sympathie (comme contagion émotionnelle) est une *qualité de la nature humaine* (T 2.1.11.2), et qu'elle est *naturelle* (EPM 7.2), il s'ensuit que la sympathie est à la fois une constitution primaire de la nature humaine, et une de ses propriétés constantes. Enfin, il faut dire que pour Hume les qualités ou les propriétés originelles sont souvent considérées comme

sympathique ne se déclenche pas, à savoir lorsque notre attention est fixée sur nous-mêmes : « [...] quand le moi fait l'objet d'une passion, il n'est pas naturel d'en abandonner la considération avant que la passion ne s'épuise » (T 2.2.2.17). En d'autres mots, lorsqu'une passion est présente en nous, elle empêche qu'on puisse ressentir ce qui affecte l'autre personne, même si elle se trouve devant nos yeux. Voilà le seul obstacle à ce principe de communication des passions. Dans toute autre situation, la sympathie nous fait immédiatement ressentir la même passion que nous croyons affecter autrui. Hume fait référence à ce type de sympathie en l'appelant aussi de la contagion émotionnelle (EPM 7.2 et NC 9-10). Cela dit, nous pouvons maintenant expliquer comment le mécanisme sympathique fonctionne afin de comprendre d'où surgissent les sentiments constituant le contenu de nos jugements moraux.

Tout d'abord, le principe sympathique opère sur l'idée d'une passion et il transforme cette idée en impression. À ce propos Hume affirme :

[a] Quand une affection s'infuse par sympathie, elle est d'abord *connue* par ses effets et par les signes extérieurs qui, dans l'attitude et la conversation, en transmettent une idée.

[b] *Cette idée se convertit sur-le-champ en une impression et elle acquiert un degré de force et de vivacité tel qu'elle se transforme dans la passion elle-même et produit une émotion égale à une affection originelle [...]*

[c] Il est évident en effet que, lorsque nous sympathisons avec les passions et les sentiments des autres, *ces mouvements apparaissent d'abord dans notre esprit comme de simples idées* et sont conçus comme appartenant à autrui, tout de même que lorsque nous concevons n'importe quelle question de fait (T 2.1.11.3/8, énumération et emphases ajoutées).

Dans [a] et [c] Hume rappelle au lecteur que la première étape dans la formulation d'une évaluation morale est l'inférence grâce à laquelle l'imagination nous porte à

étant obscures et il croit que le bon philosophe ne devrait pas essayer de les expliquer davantage (T 1.1.4.6/Intro 8).

croire que la passion *P* a motivé l'action *A*. Il affirme en effet que dans la sympathie la passion est tout d'abord connue [*known*] (non pas ressentie) et donc, qu'elle apparait dans notre esprit premièrement en tant qu'idée. En effet, comme nous avons vu dans la première section de ce chapitre, lorsque l'idée d'une passion *P* et l'idée de l'action *A* sont associées dans l'imagination, la présence de l'impression de *A* ranime l'idée de *P*. C'est cette idée de *P* qui constitue donc le point de départ du fonctionnement du mécanisme sympathique. En bref, nous ne pouvons pas sympathiser avec autrui sans avoir précédemment inféré quelle passion l'affecte (ou devrait l'affecter dans une situation donnée, voir T 2.2.9.13). Ensuite, dans [b] Hume affirme que l'idée de *P* se transforme en impression. En effet, il dit que la sympathie est la « conversion d'une idée en une impression » (T 2.3.6.8). Nous avons vu plus haut que les idées et les impressions diffèrent pour force et pour vivacité, et qu'une idée est toujours plus faible qu'une impression. Mais nous savons déjà que, selon Hume, une idée peut toujours acquérir à nouveau de la force et de la vivacité au point de devenir une impression.³⁷ Par exemple, lorsqu'à la vue d'une action nous inférons quelle passion l'a causée, l'idée de cette passion devient plus forte et vivace. Mais Hume ajoute que : « [...] outre la relation causale par laquelle nous sommes convaincus de la réalité de la passion avec laquelle nous sympathisons [...] pour sentir la sympathie dans toute sa perfection, les relations de ressemblance et de contiguïté doivent nous apporter leur concours » (T 2.1.11.8). Or, pour Hume, tous les êtres humains se ressemblent (T 2.1.11.5), mais certains se ressemblent plus que d'autres. Ainsi, quand :

³⁷ Contrairement aux autres types de principes et mécanismes psychologiques grâce auxquels les idées acquièrent de la force et de la vivacité sans devenir des impressions (voir la note 27), la sympathie permet cette transformation. C'est peut-être pour cette capacité de convertir une idée en impression que, bien que la sympathie « correspond très exactement aux opérations de notre entendement », pour Hume « elle contient même quelque chose de plus étonnant et de plus extraordinaire encore » (T 2.1.11.8).

[...] en plus de la ressemblance générale de nos natures, il existe une similitude particulière de manières, de caractères, de pays ou de langues, elle favorise la sympathie. *Plus forte est la relation entre nous et un objet, plus facile est la transition de l'imagination, qui apporte à l'idée reliée la vivacité de la conception avec laquelle nous ne cessons de former l'idée de notre propre personne* (T 2.1.11.5, emphase ajoutée).

Selon ce que Hume affirme dans cet extrait, l'idée de *P* est ravivée de façon proportionnelle à la force de la relation de ressemblance qu'il y a entre moi et la personne affectée par *P*. De plus, dans un autre passage, Hume affirme la même chose pour ce qui concerne la relation de contiguïté :

Nous sympathisons avec des personnes proches de nous plus qu'avec des personnes éloignées, avec des gens de notre connaissance plus qu'avec des inconnus, avec nos compatriotes plus qu'avec des étrangers (T 3.3.1.14).

Enfin, étant donné que, pour Hume, les liens de parentalité *dépendent* de la relation de causalité, l'imagination peut mettre en relation causale l'idée de moi avec l'idée de mon père, ou avec celle d'autres membres de ma famille.³⁸ Tout comme pour les autres deux relations, la force de la relation de causalité a des répercussions sur l'idée de *P* de manière que, plus la relation de causalité est forte, plus l'idée de *P* acquière de la force et de la vivacité. Ceci en tenant compte du fait que « chaque degré d'éloignement affaiblit considérablement la relation » (T 1.1.4.3). Ainsi, grâce à la sympathie qui, comme l'imagination, est influencée par les associations et les relations des idées, l'idée de *P* est transformée en impression.

³⁸ Dans le livre I du *Traité*, Hume semble affirmer très clairement qu'il n'y a pas une simple et constante impression du moi. Par contre, dans le contexte de l'explication du fonctionnement du mécanisme sympathique, Hume fait appel à l'impression du moi. Herdt (1997, p.41) affirme à ce propos que : « [b]ien que nous n'ayons pas une simple impression de sensation du moi, nos passions et nos soucis peuvent créer une unité dans la masse des perceptions [*bundle of perceptions*] ». De façon similaire, Pitson (2002, p. 85) semble considérer que « le caractère donne une certaine continuité à la masse ou aux systèmes de perceptions en lesquelles le soi consiste ». Selon ces deux auteurs, l'idée de soi à laquelle Hume fait référence dans le premier livre du *Traité* est différente de celle mentionnée dans les autres deux livres.

Avant d'expliquer comment ce mécanisme sympathique donne lieu aux sentiments constituant les jugements moraux, j'aimerais ajouter une précision.

Comme nous l'avons mentionné antérieurement, pour Hume nous aboutissons à croire qu'une passion *P* a causé l'action *A* à travers une inférence qui est possible puisque l'idée de *P* et l'idée de *A* sont devenue associées dans l'imagination. Lorsque nous inférons *P* en partant de l'impression de *A*, l'idée de *P* acquière de la force et de la vivacité, mais cela ne la transforme pas encore en impression (voir note 27) (flèche 1). Hume clarifie que pour que cette conversion ait lieu, l'intervention de la sympathie est nécessaire.

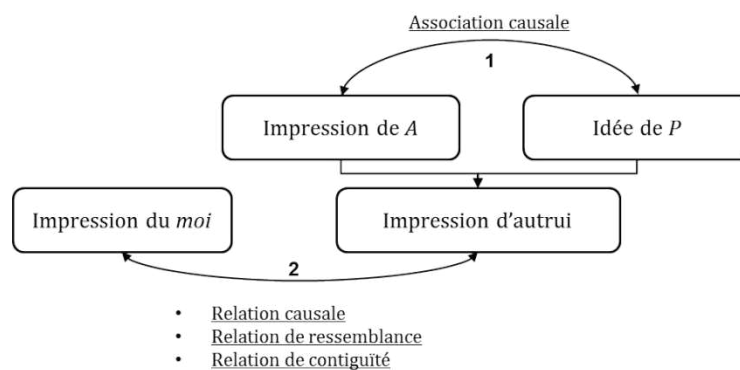


Fig. 2.1.2 L'influence des relations des idées et de leur association sur la sympathie

La *sympathie* a lieu non pas sur la base d'une association d'idées, mais sur la capacité de l'imagination de mettre en relation les idées. En effet, dans l'inférence qui nous permet de lire les intentions d'autrui, l'imagination passe de l'impression de *A* à l'idée de *P*. Par contre, dans la sympathie, l'esprit met en relation l'idée de la personne affectée et l'idée de *moi* (flèche 2) selon la ressemblance, la contiguïté et la causalité. L'idée du *moi* est toujours très vivace et forte (T 2.3.7.1), et elle communiquera un degré de vivacité et force à l'idée de la personne à qui nous ressemblons, de qui nous sommes proches ou qui fait partie de notre famille. Ensuite, étant donné que l'idée de la personne affectée par la passion et l'idée de

P sont dans une relation très forte, l'imagination passe aisément de l'une à l'autre. Comme le dit Herdt (1997, p.40) « [...] l'impression présente du moi peut prêter de la vivacité à l'idée d'une affection appartenant à quelqu'un d'autre à cause de la relation entre le moi et l'autre qui repose sur notre grande ressemblance aux autres créatures humaines ». Pour cette raison Hume peut affirmer que : « [t]outes ces relations, une fois rassemblées, conduisent l'impression ou la conscience de notre propre personne à l'idée des sentiments ou des passions des autres et nous les font concevoir de la façon la plus forte et la plus vive » (T 2.1.11.6). De cette manière, l'idée de *P* acquiert de la force et de la vivacité qui la transforment en impression.

En conclusion, l'idée de la passion est tout d'abord renforcée par l'inférence que nous formulons en partant de l'observation des actions d'autrui. Ensuite, cette même idée devient plus forte et vivace proportionnellement à la force de la relation de l'idée du *moi* avec l'impression présente de l'autre personne affectée par la passion. Ainsi, plus nous sommes proches de la personne affectée sur le plan spatio-temporel, plus nous lui ressemblons, plus nous la connaissons (T 2.2.4.2), plus l'idée de *P* acquerra de la force et de la vivacité par le mécanisme sympathique en changeant en impression et nous permettant de ressentir ce que l'autre ressent. Je peux maintenant passer à l'illustration de la façon dont, selon Hume, la sympathie donne lieu aux sentiments déterminant le contenu nos jugements moraux.

2.1.3 Le contenu des jugements moraux

Tel que nous l'avons mentionné plus haut, Hume soutient que les jugements moraux sont déterminés par des sentiments (T 3.1.1.26), autrement dit, que nous

distinguons les vices des vertus grâce aux sentiments de désapprobation et d’approbation qui surgissent de la simple considération [*contemplation*] de ces qualités (T 3.3.1.3/T 3.1.2.11).

Dans la littérature contemporaine concernant la théorie morale humienne, plusieurs discussions portent sur la nature de ces sentiments d’approbation et de désapprobation. Cet aspect de la théorie humienne n’est pas fondamental pour expliquer la manière dont, selon Hume, nous pouvons entraîner notre capacité à corriger les jugements biaisés. Il me semble toutefois important de clarifier la nature des sentiments d’approbation et de désapprobation. En effet, les jugements moraux reposent sur ces sentiments. Conséquemment, il faut expliquer quel est la nature de ces sentiments afin d’avoir une compréhension globale de la théorie morale humienne.

Certains auteurs (Korsgaard 1999, p. 11 ; Árdal 2001, p. 407) croient que ces sentiments ne sont qu’une forme plus calme des passions indirectes d’amour et de haine. Ils soutiennent cela pour deux raisons principales : 1) parce que dans un passage dans le *Traité*, Hume lui-même affirme que ces sentiments ne sont qu’une forme faible d’amour et de haine (T 3.3.5.1) ; 2) parce que les circonstances qui donnent lieu aux sentiments d’approbation et de désapprobation sont les mêmes que celles donnant lieu aux passions indirectes de l’amour et de la haine. Plus précisément, on ressent de l’amour lorsqu’une qualité de l’esprit, du corps ou d’un objet qui appartient à autrui est plaisante pour nous. Dans le cas où la qualité soit déplaisante on ressent de la haine.³⁹

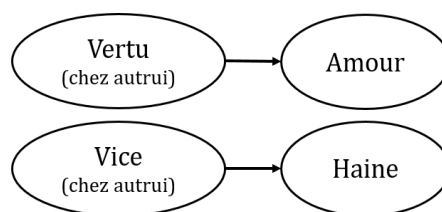


Fig. 2.1.3 Les vices, les vertus et les passions indirectes

³⁹ Par souci de clarté et de concision, je ne consacrerai pas plus d’espace à l’explication des passions indirectes. Pour plus de détails voir : T 2.1.4 et 5 aussi bien que T 2.2.1.

Les vices et les vertus sont des exemples des qualités de l'esprit (T 2.1.7) qui peuvent causer ces passions de cette manière. En effet, si la vertu est en autrui, nous ressentirons de l'amour, au contraire, si le vice est en autrui, nous ressentirons de la haine (figure 2.1.3).

Or, comme je l'ai mentionné au début de cette sous-section, les sentiments d'approbation et de désapprobations surgissent aussi de la simple considération des vertus ou des vices (en nous-mêmes ou chez autrui). Par conséquent, nous pourrions penser que si les conditions qui fournissent l'occasion aux passions indirectes sont les mêmes, il peut donc y avoir raison de penser que les sentiments d'approbation et de désapprobation sont des formes faibles des passions indirectes de l'amour ou de la haine (d'autant plus que, comme nous l'avons dit plus haut, Hume semble l'affirmer explicitement au T 3.3.5.1).

Cependant, il y a plus d'évidences textuelles en support de la thèse selon laquelle les sentiments d'approbation et de désapprobation sont un type particulier de plaisir et de douleur, que d'évidences en faveur de l'hypothèse selon laquelle elles sont des passions indirectes. Dans les paragraphes qui suivront, je vais mettre de l'avant trois raisons en faveur de l'interprétation selon laquelle les sentiments d'approbation et de désapprobation sont un type particulier de plaisir et de douleur.

Premièrement, les sentiments d'approbation et de désapprobation sont suscités par les passions que nous ressentons par sympathie. Notamment, si la passion transmise par sympathie génère un plaisir, elle sera jugée vertueuse, dans le cas contraire, elle sera un vice. Hume affirme à ce propos que :

[...] qu'est évidemment vertueuse toute qualité mentale, en nous-mêmes ou en autrui, qui nous procure une satisfaction quand nous la contemplons ou quand nous y réfléchissons, comme toute chose de nature à susciter un désagrément est vicieuse (T 3.3.1.3).

Dans ce passage, Hume affirme que l'essence même de la vertu est de procurer du plaisir, et de la douleur, dans le cas du vice (voir aussi T 2.1.7.4). Ailleurs, il dit aussi que « [n]otre approbation est *comprise* dans le plaisir immédiat », ou la douleur immédiate, que ces qualités de l'esprit d'autrui génèrent en nous (T 3.1.2.3, emphase ajoutée). Dans une troisième occurrence, il affirme que :

Le malaise et la satisfaction ne sont pas seulement inséparables du vice et de la vertu ; ils en constituent la nature et l'essence mêmes. *Approuver un caractère, c'est éprouver [feel] une jouissance lorsqu'il nous apparaît. Le désapprouver, c'est ressentir un malaise* (T 2.1.7.5, emphase ajoutée).

Dans ce passage, Hume dit que le jugement d'approbation consiste à ressentir un plaisir envers le caractère de la personne jugée. Il semble donc que l'approbation ou la désapprobation *sont* pour lui le plaisir ou le malaise surgissant de l'objet du jugement.

Deuxièmement, Hume affirme clairement que les vices et les vertus sont *la cause* des passions indirectes (T 2.1.7.2). Comme nous avons dit ci-dessus, l'essence des vices et des vertus est la douleur et le plaisir que les qualités mentales de nous-mêmes ou d'autrui causent en nous (T 2.1.7.4-5). Par contre, ces sensations de plaisir et de douleur ne peuvent pas être en même temps (1) l'essence des vices et des vertus et (2) les passions indirectes qui en dérivent (T 2.2.1.6/T 2.2.2.9/T 3.3.1.3). Par ailleurs, Hume est plus clair dans d'autres passages, où il distingue le moment où nous ressentons les sensations qui résultent de la considération des qualités mentales d'autrui, du moment où nous éprouvons la passion indirecte qui en découle (T 2.1.11.9/T 2.2.5.14/T 3.1.2.5). En conclusion, il me semble plus cohérent de soutenir que les sensations de plaisir ou de douleur causés par la considération de certaines qualités mentales en nous-mêmes ou chez autrui *sont la cause* des passions indirectes et ne leur sont pas identiques.

Troisièmement, Hume consacre deux sections entières du *Traité* (T 2.2.3/T 2.2.4) à énumérer certaines particularités des passions de l'amour et de la haine lorsqu'elles sont faibles. Si pour Hume les sentiments d'approbation et de désapprobations étaient effectivement de l'amour et de la haine affaiblies, pourquoi ne l'a-t-il pas dit explicitement dans ces mêmes sections ? Pourquoi ne l'aurait-il pas explicité dans un contexte où cela aurait été pertinent ?⁴⁰

Enfin, voici un passage où Hume souligne à plusieurs reprises que les sentiments d'approbation et de désapprobation concernant les vices et les vertus ne sont qu'un type particulier de plaisir et de douleur :

[...] puisque les impressions distinctives qui nous font connaître le bien ou le mal ne sont rien d'autre que des peines ou des plaisirs *particuliers*, il s'ensuit que dans toutes les enquêtes sur ces distinctions morales, il suffira de montrer les principes qui nous font éprouver une satisfaction ou un désagrément à considérer un caractère, pour nous assurer de la cause qui le rend louable ou blâmable. Une action, un sentiment ou un caractère sont vertueux ou vicieux : pourquoi ? Parce que *leur spectacle cause un plaisir ou un désagrément d'un genre particulier*. Donc, en donnant une raison du plaisir ou du désagrément, nous expliquons suffisamment le vice ou la vertu. Avoir le sens de la vertu n'est rien d'autre qu'*éprouver* une satisfaction d'un genre particulier à considérer un caractère. *L'impression* même constitue notre éloge ou notre admiration. Nous n'allons pas plus loin et nous ne recherchons pas la cause de la satisfaction. Nous n'inférons pas qu'un caractère est vertueux parce qu'il plaît, mais en sentant qu'il plaît d'une manière particulière, nous éprouvons effectivement qu'il est vertueux. Il en est de même que dans nos jugements sur tous les genres de beauté, sur les goûts et les sensations. *Notre approbation est comprise dans le plaisir immédiat qu'ils nous procurent* (T 3.1.2.3, deuxième et cinquième emphases ajoutées).

Dans ce passage, Hume souligne souvent le fait que les sentiments d'approbation et de désapprobation qui surgissent à la vue d'un caractère sont des plaisirs et de douleurs d'un *type particulier*. De fait, Hume prend en considération le fait que

⁴⁰ Hume affirme à ce propos que : « [...] c'est, en philosophie, une maxime inviolable : lorsqu'une cause particulière quelconque suffit pour un effet, nous devons nous en satisfaire et ne devons pas multiplier les causes sans nécessité » (T 3.3.1.10).

nous pouvons ressentir plusieurs types de plaisirs et de douleurs, et il affirme que chacune de ces sensations est phénoménologiquement distincte des autres (Debes 2012, p. 94), dans la mesure où elles sont perçues comme étant différentes des autres. Il est vrai qu'on ne mélange pas la sensation plaisante qui nous donne un bon vin, avec le plaisir d'un bon bain chaud en hiver. De la même façon, à son avis, nous sommes capables de discerner entre ces deux types de sensations et les sensations causées par les vices et les vertus (T 3.1.2.4). Notamment, Hume tient à souligner que le sentiment d'approbation (ou de désapprobation) qu'on ressent à la vue d'une maison belle (ou laide) d'un de mes bons amis est bien distinct des sentiments d'approbation (ou de désapprobation) qu'on peut ressentir en contemplant la vertu (ou le vice) de son caractère (T 3.3.5.6).

Ces sentiments sont d'une certaine manière « radicalement différents des autres sentiments » (Debes 2012, p. 91). Cette différence a une fonction épistémique, puisque si ce n'était de celle-ci, nous serions incapables de distinguer les vices des vertus (Debes 2012, p. 110).

Eu égard à ce qui précède, je suis d'accord avec Debes lorsqu'il affirme que la réflexion de Hume au T 3.3.5.1 ne suffit pas pour soutenir que les sentiments d'approbation et de désapprobation sont une forme affaiblie d'amour et de haine (Debes 2012, p. 109). Conséquemment, je pense que l'interprétation selon laquelle les sentiments d'approbation et de désapprobation sont des types de sensation particulière est la meilleure.

Avant d'illustrer dans quels cas, selon Hume, nos jugements moraux peuvent être biaisés, j'aimerais ajouter une dernière remarque. Pour Hume, les êtres humains ont un sens moral en tant qu'ils ont la capacité de ressentir de l'approbation et de la désapprobation envers un caractère (Korsgaard 1999, p. 13). Cette faculté est

un principe constant de la nature humaine, mais elle peut être défectueuse, par exemple si l'esprit est en proie à une maladie ou à la folie (T 3.1.2.8).

2.2 La sympathie limitée et les erreurs de jugement

Comme nous avons vu, dans la citation au tout début de ce chapitre, Hume affirme qu'il faut raisonner beaucoup et correctement afin de ressentir *le sentiment moral approprié*. Plus précisément, Hume affirme qu'il faut premièrement « tirer les justes conclusions » (EMP 1.9) afin d'éprouver ce *sentiment approprié*. Or, comme nous l'avons vu, la première *conclusion* qu'on tire lorsqu'on évalue moralement a comme objet la motivation de l'agent et elle dépend des croyances que nous avons accumulées pendant notre vie, grâce à nos expériences et par l'éducation. C'est pourquoi je crois que pour Hume le premier faux pas que nous pouvons commettre par le biais de la raison consiste à tirer des mauvaises conclusions en se laissant conduire les yeux fermés par nos croyances.

2.2.1 Percevoir les passions d'autrui

J'ai expliqué dans la première section de ce chapitre comment l'expérience et l'habitude provoquent l'association causale de l'idée d'une passion à l'idée d'une action et comment, par la suite, la raison peut inférer quelle passion a causé l'action à la vue de cette dernière.

L'association entre l'idée d'une passion et l'idée d'une action dans l'imagination dérive de l'expérience constante que nous faisons de la cooccurrence de leurs impressions. Notamment, si chaque fois que je suis nerveuse, je mange du chocolat, il semble tout à fait justifiable pour moi d'inférer que j'aurais envie de chocolat la prochaine fois que je serai nerveuse. Je dis que cette inférence est

justifiable dans le sens où *je fais expérience* de la passion et de l'action dont les idées deviennent associées dans l'imagination permettant ensuite à la raison de formuler des inférences. En effet, pour Hume, nos pensées et nos sentiments nous sont toujours connus.⁴¹ Ainsi, il affirme que le moi est « cette personne identique, dont notre conscience pénètre intimement les pensées, les actions et les sensations [...] » (T 2.2.1.2, voir aussi T 2.1.2.2 et T 2.2.2.15 sur la même idée). Il semblerait dès lors que pour Hume nous ne pouvons pas nous tromper en « lisant » notre propre esprit.

Au contraire, lorsque nous jugeons moralement les passions qui ont motivé l'action d'autrui, il surgit un problème : Hume affirme que « [c]hez autrui, nulle passion ne se découvre immédiatement à l'esprit. Nous ne sommes conscients que de ses causes et de ses effets » (T 3.3.1.7). Il est clair, en effet, qu'on ne peut pas appréhender les passions de quelqu'un de la même manière que nous pouvons saisir les nôtres. En effet, nous ne pouvons pas « voir » les passions qui affectent autrui et il semble bien étrange d'affirmer que je puisse faire expérience des passions qui affectent autrui.

Il est cependant nécessaire que l'on conçoive les passions des autres d'une certaine façon afin de pouvoir inférer celles qui sont la cause des leurs actions et pour les juger moralement. Voici une solution plausible à cette impasse, et qui serait au surplus compatible avec le système philosophique humien. Nous pourrions penser que l'association entre l'idée de l'action *A* et l'idée de la passion *P* découle toujours de notre propre expérience, qu'ensuite nous étendons aux autres.⁴² Il serait possible d'inférer la passion *P* à partir de l'action *A* en se basant

⁴¹ Pour plus des détails sur l'idée selon laquelle nos pensées nous sont connues voir : Waxman, Wayne. 1994. *Hume's Theory of Consciousness*. London ; Cambridge University Press, pp.203-239.

⁴² Je suis d'accord avec l'interprétation de Debes sur ce point (Debes 2007a, p. 316). Je soutiens ici le même point que Anik Waldow (2008, pp. 63) qui ajoute à ce propos que : « Of course, this is

sur l'association entre l'idée de *P* et l'idée de *A* que notre esprit a établie suite à notre propre expérience. D'autant plus que, selon Hume :

[I]l est universellement reconnu qu'il y a une grande uniformité entre les actions des hommes, dans toutes les nations et à toutes les époques, et que la nature humaine demeure toujours la même dans ses principes et ses opérations. *Les mêmes motifs produisent toujours les mêmes actions, les mêmes événements s'ensuivent des mêmes causes* (EEH 8.1.7, emphase ajoutée).

Ce passage nous permettrait aussi d'affirmer que ce changement de perspective serait justifié par le fait que *les mêmes motifs produisent toujours les mêmes actions*, peu importe la culture d'appartenance de l'agent.

Toutefois, ailleurs dans le *Traité* aussi bien que dans d'autres textes, Hume tient compte de l'influence que l'éducation et la vie en société (EEH 8.10, voir aussi ENG 1) ont sur nos comportements. Il affirme : 1) que la ressemblance entre les habitants d'un même pays dépend exactement du fait qu'ils habitent dans le même pays (T 2.1.11.2); 2) que les préjugés sont acquis à travers l'éducation sous le même gouvernement (ENG 1); 3) que l'éducation a le même effet que l'association des idées sur l'imagination puisqu'elle renforce certaines idées en générant ainsi des croyances; 4) que « les différentes conditions de vie influencent toute la structure, externe et interne; et ces différentes conditions proviennent nécessairement, car uniformément, des principes eux-mêmes nécessaires et uniformes de la nature humaine » (T 2.3.1.9); 5) que chaque individu change beaucoup pendant sa vie (EEH 8.10-11).

Toutes ces observations mènent à la même conclusion : Hume veut ici soutenir que, bien qu'il y ait une certaine ressemblance entre les êtres humains, la façon propre à chacun d'exprimer ses passions peut changer, d'une part par rapport à

not to deny that these ideas can represent passions which we have never felt ourselves: to understand that you are in pain and to feel for you does not require me to be familiar with the precise pain you are enduring. All that is needed for my sympathy with you is that I have once felt some sort of pain myself, for then there is an idea available by which I can represent and understand your pain ».

lui-même (puisque'il change dans le temps), d'autre part, par rapport aux habitants d'autres pays (qui ont une culture différente). Bref, Hume soutient qu'il y a une certaine uniformité dans les êtres humains en acceptant qu'il y ait aussi une certaine diversité. Pour cette raison, la probabilité d'erreur est forte en étendant nos croyances concernant le lien causal entre une passion et une action à quelqu'un d'autre afin de « lire » ses intentions, surtout si l'individu qu'on essaie de comprendre est très différent de nous.⁴³

En conclusion, il n'y a aucune évidence textuelle qui confirme que selon Hume la nature humaine est uniforme à un point tel que nous ne tombons pas en erreur en inférant la passion qui affecte autrui à la vue de son action en se basant sur notre propre expérience. En outre, il n'est pas évident si pour Hume il y a certaines passions qui, de par la nature humaine, causent universellement les mêmes actions. Puisque ce n'est pas le but de ce texte d'explorer cette possibilité et de répondre à cette question, je ne pourrai poursuivre davantage sur ce sujet.

Eu égard de ce qui précède, il semblerait donc qu'une première erreur que nous pouvons faire en formulant un jugement moral consiste à mésinterpréter l'action de quelqu'un d'autre sur la base d'une association entre l'idée d'une passion et l'idée d'une action qui découle de notre expérience particulière. Par ailleurs, notre expérience peut être enrichie en côtoyant des personnes qui expriment leurs passions différemment et qui nous communiquent leurs états d'esprit (Baier & Waldow 2008, p. 64). De telles interactions sociales nous permettraient de sortir de notre perspective limitée.

⁴³ Pour plus de détails à ce propos voir surtout T 2.3.1.12 et EEH 8.

2.2.2 Le caractère

Dans les passages que j'ai cités jusqu'à maintenant, Hume affirme que nous jugeons les motivations à l'action plutôt que l'action en elle-même, et que les motivations sont toujours des passions. Toutefois, dans d'autres passages du *Traité*, Hume soutient qu'on évalue du vice et de la vertu en considérant plutôt le *caractère* d'un agent (T 3.3.1.5). Pour l'auteur, le caractère est un ensemble stable de passions qui affectent régulièrement le même sujet. En d'autres mots, lorsqu'on parle du caractère de quelqu'un Hume n'indique rien d'autre que les passions qui se présentent *avec constance* dans son esprit (T 3.3.1.4).

Or, nous pouvons découvrir les passions qui apparaissent constamment dans l'esprit de quelqu'un par le biais de l'expérience, mais aussi par celui de l'éducation. Dans le premier cas, la constance dans les actions d'autrui est l'indice de son caractère. Si la constance est telle que nous inférons plusieurs fois que l'action *A* est motivée par la passion *P*, nous aboutissons à croire que la passion *P* est la cause de *A* chez le sujet et elle sera une des passions constantes qui font partie de son caractère. Dans le deuxième cas, comme nous avons mentionné, l'éducation a le même effet que l'habitude sur nos idées et peut générer des croyances (T 1.3.9.17).⁴⁴ Ceci étant dit, Hume affirme :

Qu'une *action* soit vertueuse ou vicieuse, elle ne l'est qu'en tant que signe d'une certaine qualité ou d'un certain caractère. Elle doit [must] dépendre de principes durables de l'esprit qui s'étendent à la conduite entière et participent au caractère de la personne (T 3.3.1.4, emphase ajoutée).

Dans ce passage, Hume soutient qu'une évaluation morale a toujours comme objet le caractère de la personne. Toutefois, selon une telle lecture de ce passage, pour Hume nous ne jugerions que les passions stables, et donc, nous n'évaluerions

⁴⁴ Dans ce cas, nous aboutissons à penser qu'une passion a été la cause d'une action, ou aussi qu'une passion est un trait de caractère de quelqu'un, puisque son idée a acquis de la force et de la vivacité grâce à l'éducation.

moralement que des sujets que nous connaissons (ou que nous croyons connaître).⁴⁵

Par contre, nous remarquons bien souvent le contraire, à savoir qu'on juge aussi les individus qu'on ne connaît pas (ou que nous savons ne pas connaître). En outre, il nous arrive aussi de juger moralement quelqu'un dont nous connaissons très bien le caractère lorsqu'il est motivé par des passions qui ne sont pas en lien avec son caractère. Il me semble peu charitable de penser que Hume ait pu négliger de tenir en compte d'un fait aussi flagrant de notre vie morale. Je crois plutôt que dans le passage cité auparavant, Hume ait plutôt voulu dire que les bons jugements moraux doivent porter sur le caractère de quelqu'un et non pas sur une passion passagère. En d'autres mots, il soutient que les bonnes évaluations morales portent seulement sur le caractère des individus et non pas les passions qui des temps en temps les poussent à agir.⁴⁶

On pourrait ici m'accuser de lire trop dans ce passage, mais j'estime que le propos de Hume est encore plus évident dans le passage suivant, où il affirme que :

Dans nos recherches sur l'origine de la morale, nous ne devons pas considérer une action en elle-même, quelle qu'elle soit, mais seulement la qualité ou le caractère d'où elle procède. Eux seuls sont assez durables pour toucher les sentiments que nous avons à l'égard d'une personne (T 3.3.1.5).

Suite à cet autre passage, il est au moins plausible de penser que pour Hume juger des passions passagères plutôt que du caractère de quelqu'un nous mènerait à erreur dans notre évaluation morale. Je crois donc que, selon Hume, nos

⁴⁵ Par exemple lorsqu'on suit des règles générales nous jugeons sur la base des préjugés T.1.3.13.18 et T.2.1.6.8.

⁴⁶ En effet, le terme anglais que Hume utilise dans le passage cité est « *must* ». « *must*, n.3 and adj.3. » *OED Online*. Oxford University Press, March 2014. Web. 31 March 2014, Def.1: « A use of 'must' (MUST v.¹) to express a command, obligation, or necessity; (hence) an obligation, a duty; a compulsion ».

évaluations morales ont généralement comme objet autant les passions passagères que celles plus stables constituant le caractère. Mais aussi que nous ne ressentirons pas *le sentiment approprié* en contemplant une passion passagère qui a occasionnellement provoqué une action, plutôt qu'un caractère.⁴⁷

En conclusion, dans cette section nous avons vu que, selon Hume, l'objet du jugement moral est l'intention qui cause les actions d'un individu. En outre, comme j'ai dit plus haut, pour Hume, étendre notre expérience personnelle concernant le lien entre une passion et une action afin d'inférer quelle passion a motivé l'action chez autrui pourrait nous induire en erreur et éventuellement nous empêcher de ressentir le *sentiment approprié*. De la même manière, juger des passions passagères plutôt que des passions stables chez quelqu'un entraverait notre capacité à ressentir ce *sentiment approprié*. Au contraire, en évaluant le caractère de l'agent nous pouvons, conformément à la théorie humienne, éviter ces deux erreurs. En effet, pour juger du caractère de quelqu'un il faut connaître la personne dans ses détails et cela nous permet aussi d'éviter de mésinterpréter les signes de ses passions. On pourrait maintenant se demander c'est quoi qui nous motiverait à regarder au caractère de quelqu'un, plutôt qu'à ses passions passagères, lorsqu'on formule notre évaluation morale. Nous verrons cela dans la troisième partie de ce chapitre. Dans la prochaine sous-section, je présenterai la dernière cause qui nous mène à ressentir des sentiments inappropriés, et ce, même lorsqu'on a bien raisonné concernant le caractère du sujet que nous jugeons.

⁴⁷ Comme mentionné au début de ce chapitre, pour Hume une évaluation morale ne peut pas être erronée. Lorsque j'affirme le contraire, je ne m'oppose donc pas à une lecture standard de Hume, je veux plutôt dire que pour l'auteur le sentiment qui constitue le contenu d'une évaluation morale peut être inapproprié (p.ex. lorsqu'il s'agit de l'évaluation d'une passion passagère plutôt que du caractère de quelqu'un).

2.2.3 Les « erreurs » de la sympathie limitée

Au tout début de ce chapitre, j'ai cité un passage de *l'Enquête sur les Principes de la Morale* dans lequel Hume soutient que les distinctions morales découlent en partie d'un raisonnement. Ce raisonnement, lorsque correctement employé, ouvre la voie aux sentiments appropriés qui déterminent au final le contenu des jugements moraux. Comme nous l'avons vu, les sentiments d'approbation et de désapprobation constituent le contenu de nos évaluations morales et ils se manifestent une fois que nous arrivons à ressentir la passion qui affecte autrui, grâce au mécanisme sympathique. Il faudrait maintenant comprendre si ces sentiments sont *les sentiments appropriés* que Hume mentionne dans le passage cité. Nous verrons dans cette sous-section pourquoi ce n'est pas le cas.

Tout d'abord, il est évident que le fonctionnement du mécanisme sympathique tel qu'il a été illustré précédemment est imbibé de subjectivité.⁴⁸ En effet, ce principe universel de la nature humaine fait en sorte que nous ressentons la même passion qui a causé l'action de quelqu'un d'autre avec une force proportionnelle à la ressemblance (T 3.3.3.4) et à la proximité que nous avons au sujet avec qui nous sympathisons. De plus, la force de la passion est encore plus frappante si la personne qui en est affectée est liée à nous par une relation de parentèle (T 1.1.4.3). Il n'est pas nécessaire de préciser davantage cette question : il est clair que le mécanisme sympathique est influencé par nos liens personnels. Étant donné que cette sympathie trouve son origine dans le point de vue propre à la personne qui sympathise, Hume la considère comme *limitée* (ou *partielle*) (T 2.2.7.5/T 3.3.1.23). De cette *sympathie limitée*, il s'ensuit que nous sommes plus prompts à

⁴⁸ Debes (2007a, p.317) supporte cette même lecture du texte. Il affirme: « [...] it is precisely this essential relation back to the self, or *self-association*, in the sympathetic process that provokes part of the problem of subjective bias in our moral sentiments [...] ».

approuver les vertus de ceux et celles qui nous sont proches dans l'espace et dans le temps, que les vertus de ceux et celles qui sont loin de nous. Pour la même raison, nous approuvons plus facilement les caractères qui sont le plus semblables au nôtre (T 3.3.3.4). Qui plus est, selon Hume, le fait qu'on sympathise de façon limitée est non seulement une qualité constante de notre esprit comme nous l'avons expliqué auparavant, mais aussi une qualité originelle, dans le sens qu'elle fait partie de la nature humaine (voir note 34). Hume dit notamment que le fait que « [...] notre attention la plus forte est réservée à notre famille et à nos relations et ce n'est que la plus faible qui se porte jusqu'aux personnes qui nous sont étrangères et indifférentes » est une partie de la « structure originelle de notre esprit » (T 3.2.2.8).

En outre, cette sympathie est influencée par les croyances qui sont le fruit de l'expérience de même que par les croyances assimilées par l'éducation et la vie en société (qui sont plus précisément des « préjugés »). Hume affirme notamment qu'une personne qui est influencée par les préjugés « [...] garde avec obstination sa position naturelle » (ENG 21). Il me semble plausible de penser que cette *position naturelle* dont Hume parle ici coïncide avec le point de vue particulier qui est occupé de manière constante par chacun d'entre nous (voir note 34). En d'autres mots, si nous ne nous sortons pas de notre point de vue particulier, nous ne nous libérerons pas de chaque préjugé qui, selon Hume, est « destructeur du jugement sain » (ENG 22).

En conclusion, Hume affirme explicitement que la partialité qui caractérise la sympathie limitée est « vicieuse et immorale » (T 3.2.2.8). Ainsi, elle fait obstacle au *sentiment approprié* et par la suite au *bon jugement*. Je consacrerai la prochaine

section à l'explication de comment la façon dont, selon lui, nous pouvons corriger les jugements moraux qui sont le produit d'une sympathie limitée.

2.3 La correction des sentiments inappropriés

Avant d'illustrer comment, selon Hume, nous pouvons corriger les jugements moraux fautifs, j'aimerais récapituler certains éléments de la discussion que nous avons eue jusqu'à maintenant.

Dans la première section de ce chapitre, nous avons vu que pour Hume nos jugements portent sur les intentions et non pas sur l'action en elle-même, et que nous pouvons saisir les intentions d'autrui grâce à des raisonnements causaux (ce qu'il appelle des « inférences »). Le résultat d'un raisonnement causal est l'idée de la passion qui anime autrui, que nous finissons par ressentir grâce au mécanisme sympathique. Ce mécanisme, tel que Hume le décrit pour la première fois au T 2.1.11, prend la forme d'une contagion émotionnelle. Enfin, la passion qu'on a « attrapée » par sympathie provoque en nous des sentiments d'approbation ou de désapprobation, dépendamment du fait que la sensation qu'elle génère en nous est plaisante ou déplaisante.

Dans la deuxième section, j'ai montré quel genre d'erreurs nous portent à ressentir les sentiments inappropriés. Premièrement, j'ai expliqué qu'il est possible de commettre des erreurs lorsqu'on infère le caractère de quelqu'un à partir de son action. Par exemple, nous pouvons nous tromper en assumant que certains comportements qui chez nous sont un signe d'envie sont le signe de la même passion chez autrui. D'ailleurs, l'inférence qui nous permet de conclure que

l'envie a motivé l'action peut reposer sur des croyances acquises par l'éducation, mais qui se révèlent fausses ou inappropriées pour la situation donnée. Deuxièmement, j'ai soutenu que pour Hume nous pouvons ressentir les sentiments appropriés seulement lorsqu'on juge du caractère de quelqu'un plutôt que de ses passions passagères. Enfin, j'ai voulu mettre en relief que la sympathie comme contagion émotionnelle (la sympathie limitée) peut, elle aussi, donner lieu à des erreurs. En effet, la sympathie limitée nous mène à ressentir les passions des autres de notre point de vue particulier ; ce dernier est partiel, puisqu'il est influencé par les relations personnelles que nous entretenons avec la personne que nous sommes en train de juger. C'est cette partialité de la sympathie limitée qui rend le jugement biaisé. De cette façon, autant nos inférences que la sympathie limitée peuvent nous induire en erreur dans la formulation des jugements moraux. Mais Hume ne se contente pas d'affirmer que nos jugements moraux puissent être erronés : il offre une solution afin d'éviter ces erreurs, à savoir, l'adoption d'un *point de vue commun* (dorénavant PVC) lorsqu'on sympathise avec un agent. Je crois que c'est en partie à cause de la subjectivité qui affecte le fonctionnement associatif de la sympathie limitée que Hume fait appel au point de vue commun pour s'assurer qu'on formule des jugements moraux de façon non biaisée.⁴⁹

La prochaine section portera sur la nature du PVC. Nous verrons ce en quoi il consiste, quelles sont les motivations pour l'adopter et quelles sont les conséquences du changement de perspective sur les sentiments d'approbation et de désapprobation. J'argumenterai brièvement pour montrer qu'occuper le point de vue commun ne serait pas utile pour réviser le jugement moral si cela n'entraînait pas aussi une remise en question de nos raisonnements concernant les intentions des agents. De plus, nous verrons que considérer un caractère à partir

⁴⁹ Sur cette question, voir aussi Debes 2007a, p. 318.

du PVC nous permet, selon Hume, de ressentir *le sentiment approprié* auquel il fait référence dans le passage de l'*Enquête sur les Principes de la Morale* que nous avons cité au tout début de ce chapitre. Enfin, je montrerai que, pour Hume, c'est grâce à l'expérience et à la pratique que nous pouvons apprendre à occuper automatiquement le PVC.

2.3.1 Pourquoi atteindre le PVC ?

Dans le *Traité*, Hume met de l'avant deux raisons pour lesquelles nous changeons la perspective d'où nous jugeons les individus : (1) afin de ne pas être en désaccord avec les autres et (2) afin d'éviter que nos jugements soient incohérents.

Premièrement, nous finissons par juger une situation donnée à partir du PVC afin d'éviter de ressentir le malaise qui découlerait d'un désaccord avec autrui. En effet, tel que je l'ai mentionné plus haut, la sympathie limitée consiste à juger une situation à partir de notre propre point de vue. Or, selon Hume, tant que la sympathie limitée nous guide dans la formulation de jugements, ce dernier sera largement influencé par nos préférences personnelles et nos biais. Ainsi, si nous jugeons en demeurant dans notre position particulière, il est très probable que nous soyons en désaccord les uns avec les autres. Hume affirme à ce sujet que :

[...] il est impossible que nous [ne] puissions jamais converser ensemble en des termes raisonnables, si chacun de nous considère les personnes et les caractères uniquement tels qu'ils apparaissent de son point de vue singulier (T 3.3.1.15).

Bref, c'est pour communiquer paisiblement avec les autres membres de notre société que nous corrigeons les évaluations morales qui suivent d'une sympathie

limitée. Or, comme nous l'avons vu, pour Hume nous sommes motivés à agir seulement par des désirs ou par des aversions : nous poursuivons ce qui est plaisant et nous fuyons ce qui est déplaisant. Puisque, selon la théorie humienne, être en désaccord avec les autres est une source de déplaisir, il est normal d'essayer de l'éviter. Hume affirme à ce sujet que nous ne pourrions jamais être en accord les uns avec les autres, à moins d'occuper un point de vue à partir duquel le caractère et les attitudes de l'agent sont perçues de la même manière par tous (T 3.3.1.30). Ce point de vue est le PVC.⁵⁰

Deuxièmement, bien que Hume ne le dise pas explicitement, je crois qu'il accepterait de dire que nous occupons le PVC afin d'éviter que nos jugements sur autrui soient incohérents ou contradictoires.⁵¹ En effet, Hume affirme que :

[...] afin de prévenir ces *contradictions* continuelles et de parvenir à un jugement des choses qui soit plus *stable*, nous choisissons des points de vue *fermes* et *généraux* et, dans nos pensées, nous nous y plaçons toujours, quelle que puisse être notre situation présente (T 3.3.1.15).

Il nous arrive en effet de changer d'avis sur une personne, parfois même radicalement, et Hume affirme clairement que l'inconstance et l'instabilité dans une idée cause en nous un certain malaise [*uneasiness*] (T 2.3.10.12). Or, cette indécision dans le jugement peut être due à une variation dans nos croyances aussi bien qu'à un changement des circonstances dans lesquelles nous nous trouvons.

D'une part nos croyances peuvent changer au cours du temps. Comme nous l'avons maintes fois répété, ces croyances influencent les raisonnements qui nous permettent de saisir les intentions d'autrui qui, à leur tour, constituent le point de

⁵⁰ Voir aussi Cohon (1997) et Korsgaard (1999).

⁵¹ À l'époque, le terme « contradiction » était employé dans le sens de « opposition ».

départ de l'évaluation morale. Ainsi, si nos croyances concernant le caractère d'une personne changent, il en sera de même pour nos jugements moraux.

D'autre part, Hume tient compte du fait que la force des relations qui lient le *moi* avec les autres peuvent changer dans le cours du temps. Il dit à ce propos qu'en effet « [n]otre situation à l'égard des personnes et des objets varie continuellement, et il se peut qu'un homme qui se trouve loin de nous devienne, en peu de temps, une relation familière » (T 3.3.1.15). Mais, comme nous l'avons vu, le fonctionnement de la sympathie dépend de la force des relations de contiguïté, de causalité et de ressemblance entre le *moi* et les autres. Conséquemment, chaque augmentation ou diminution d'intensité de ces relations modifie la force de la passion communiquée par sympathie.⁵²

En somme, l'inconstance de nos jugements moraux est déplaisante pour nous tout comme l'est le fait d'être en désaccord avec les autres. Ainsi, si la solution proposée par Hume au malaise qui découle d'une dissension entre moi et autrui est celle d'occuper le PVC, je crois qu'il est au moins plausible de penser qu'il aurait suggéré la même solution pour le malaise qui dérive d'une incohérence avec nous-mêmes. Ceci étant dit, je peux maintenant expliquer ce qui caractérise le PVC.

⁵² Hume affirme que nous occupons le PVC pour éviter des contradictions tout de suite après avoir dit que nous ne pourrions pas converser les uns avec les autres en demeurant au sein de notre point de vue particulier. À cause de l'ordre dans lequel ces idées sont présentées, nous pourrions penser que selon Hume la contradiction n'est pas avec nous-mêmes, mais plutôt avec les autres. Pourtant, comme le soutient Korsgaard (1999, p. 14), il est difficile de comprendre pourquoi Hume parlerait de contradiction. Dans quel cas en effet pourrait-on être en contradiction avec autrui ? Bref, le terme contradiction ne semble pas approprié pour décrire le désaccord qu'on peut avoir avec les autres. Au contraire, nous pouvons être en contradiction avec nous-mêmes et c'est pour fuir le malaise qui en découle que nous nous plaçons dans le PVC. Il est aussi vrai, comme le dit Korsgaard, que « [...] dans tous les cas, nous ne nous attendons pas de nous réjouir des mêmes choses à tout moment [*at all times*] » (Korsgaard 1999, p. 14). Notamment, il semble normal que dans notre âge adulte nous puissions changer d'avis sur les personnes qu'on côtoyait pendant notre adolescence. Cela n'empêche pas que ressentions un malaise lorsque nous ne comprenons pas le caractère de quelqu'un ou lorsque nos opinions là-dessus changent constamment dans un espace de temps assez court. Korsgaard (1999) interprète différemment l'emploi du mot contradiction dans le passage cité.

2.3.2 Le PVC

Dans le troisième et dernier livre du *Traité*, Hume affirme que :

Le seul point de vue en lequel nos sentiments coïncident avec ceux des autres, c'est lorsque nous considérons la tendance d'une passion à être avantageuse ou dommageable pour ceux qui ont un rapport direct ou une relation immédiate avec celui qui la possède (T 3.3.3.2, emphase ajoutée).

Considérons les conséquences de ce que Hume dit dans le passage cité.

Premièrement, Hume affirme qu'il y a un point de vue qui nous permet d'avoir les mêmes sentiments (d'approbation ou de désapprobation) qu'autrui. Compte tenu du fait que nos jugements découlent au final de nos sentiments, ressentir les mêmes sentiments qu'autrui implique de partager leurs jugements. J'ai précédemment montré que pour Hume nous devons occuper le PVC pour être en accord avec autrui. Le point de vue dont il parle dans ce passage correspond donc au PVC dont nous venons de parler.

Deuxièmement, Hume affirme que lorsque nous considérons les avantages et les désavantages d'un caractère donné *pour son entourage*, c'est-à-dire, lorsque nous adoptons le PVC, nous sommes tous d'accord les uns avec les autres. L'entourage de l'agent consiste en l'ensemble des personnes qui sont en relation directe avec lui. Pourtant, la notion de *relation directe* est tout de même assez floue, surtout lorsqu'on se rappelle que les relations humaines sont instables et que leur force est sujette à gradation. Dès lors, nous pourrions nous demander à quel point un individu doit être proche, doit ressembler ou doit être en lien de parenté avec quelqu'un pour être en relation directe avec lui. Cela entraîne une confusion lorsque nous souhaitons occuper le PVC puisqu'on ne sait pas avec précision quel est le point de vue que nous devons prendre en considération dans la réflexion qui corrige nos jugements moraux biaisés. D'ailleurs, Hume s'exprime différemment en d'autres occasions. Par exemple il affirme que :

[...] le seul intérêt, le seul plaisir qui paraît identique à chaque spectateur *est celui de la personne dont on examine le caractère*, ou celui des personnes qui sont en lien avec elle (T 3.3.1.30, emphase ajoutée).

Dans ce cas, l'auteur semble dire que même en occupant la perspective de la personne qui est objet du jugement, nous serons tous d'accord. Finalement, il affirme ailleurs que nous portons notre attention sur l'intérêt de la société (T 3.3.1.11) ou même de l'humanité (T 3.2.5.12) lorsqu'on occupe le PVC. Bref, Hume n'est pas précis quant à la nature du PVC. Malgré cela, il n'est pas important d'analyser davantage cet aspect de la théorie humienne. En effet, je fais ici appel à la théorie morale de Hume afin d'expliquer comment, selon lui, nous pouvons corriger nos jugements moraux. Pour ce faire, il suffira se souvenir que pour Hume *la correction des jugements moraux dépend d'un changement de perspective sur la situation qu'on veut juger*.

Enfin, je crois que ce changement de perspective entraîne une variation dans la première inférence grâce à laquelle nous obtenons l'idée du caractère d'autrui. De fait, comme nous l'avons mentionné dans la deuxième section de ce chapitre, le point de vue particulier et subjectif de chacun d'entre nous est imprégné de préjugés et de croyances. Ainsi, lorsque nous sommes poussés à « transcender notre propre perspective » (Baier et Waldow 2008, p. 62-63) pour occuper le PVC, nous sommes obligés de remettre en question nos conclusions concernant le caractère d'autrui, si nous voulons effectivement comprendre comment ce dernier apparaît à son entourage, à soi-même ou à la société entière. Je ne veux pas dire ici que, lorsque nous essayons d'occuper le PVC nous abandonnons automatiquement nos croyances ou nos préjugés. Je crois plutôt que, si nous réussissons nous avons probablement dû distinguer les croyances qui étaient pertinentes pour comprendre le caractère de l'individu en question des celles qui ne l'étaient pas. En somme, je crois qu'il est cohérent avec la théorie psychologique humienne de dire que lorsque nous raisonnons sur le caractère d'un

agent à partir d'une perspective différente de la nôtre, cela nous encourage à remettre en question les croyances d'où ces inférences découlent.

2.3.3 Le PVC et la sympathie étendue

Il est clair que, pour Hume, l'aversion que nous avons à être en désaccord avec autrui ou à être incohérents nous pousse à nous placer dans le PVC et à considérer le caractère d'un agent dans l'intérêt d'autrui. Or, pour Hume, c'est la *sympathie étendue* qui nous permet de porter attention au bien-être d'autrui plutôt qu'au nôtre.⁵³ La sympathie étendue repose sur les mêmes principes de la sympathie limitée (qui sont les principes de l'imagination). En effet, la sympathie étendue, comme la sympathie limitée, consiste en une conversion d'une idée en une impression. Mais le fonctionnement de la sympathie limitée part d'une perspective intéressée, tandis que la sympathie étendue nous pousse à avoir à cœur le bien de la personne avec qui nous sympathisons. Conséquemment, lorsque nous changeons de point de vue pour juger le caractère d'autrui nous sympathisons différemment avec lui.

⁵³ Hume fait mention de deux types de sympathie étendue, comme le remarque aussi Cunningham (2004). La sympathie que Hume présente au T 2.2.9 est étendue puisque la force de la passion communiquée par sympathie nous pousse à tenir en compte non seulement la condition présente du sujet avec qui nous sympathisons, mais aussi son passé et son futur. Par conséquent, la condition du sujet finit par nous tenir à cœur. Cette forme de sympathie ne nous pousse pas à changer de point de vue, elle est donc encore affectée par les liens personnels que nous avons avec l'individu. Par contre, la sympathie présentée par Hume dans le livre III du *Traité* la sympathie est étendue en tant que dirigée vers tous les membres de la société subissant les conséquences du caractère de quelqu'un (donc, plus précisément, l'entourage de la personne en question). En outre, cette sympathie est impliquée dans le changement de point de vue, qui consiste en effet à porter attention à l'intérêt de la communauté plutôt qu'à l'intérêt personnel. Nonobstant ces différences, il semble que les deux types de sympathie ont une chose en commun : l'intérêt dans le bien d'autrui plutôt que de soi-même.

Tel que je l'ai mentionné dans la sous-section précédente, lorsqu'on change de point de vue sur une situation, la conclusion de nos inférences peut changer à son tour. Ainsi, la passion P inférée pourrait ne pas être la même qu'on aurait inférée à partir de notre point de vue particulier. Nommons cette autre passion P_2 (qui, je souligne, pourrait toujours être P).

De plus, en nous déplaçant de point de vue, nous modifions notre relation avec l'agent. En effet, Hume affirme que « l'objet a beau rester le même, si nous changeons de point de vue sur lui, nous modifions radicalement son rapport avec nous-mêmes » (T 2.2.10.3). Comme j'ai déjà expliqué, la sympathie est influencée par ces relations. En effet, l'idée d'une passion acquiert de la vivacité et de la force proportionnellement à la force des relations qu'il y a entre l'idée du *moi* et l'idée de l'agent. Ainsi, supposons que de notre point de vue les relations de contigüité (flèche 1), de causalité (flèche 2) et de ressemblance (flèche 3) ne sont pas très fortes. Conséquemment, lorsqu'on occupe notre propre point de vue, l'idée de P ne sera pas assez revigorée pour devenir une impression en elle-même.

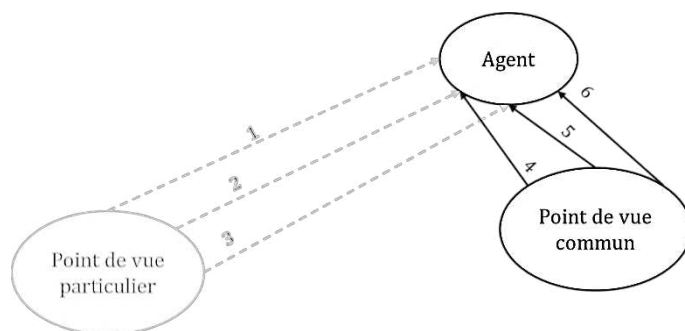


Fig. 2.3.3 La sympathie étendue implique une variation des relations entre les idées

Considérons le caractère du même agent à partir du PVC. Ce changement de perspective nous porte à imaginer que le moi soit dans des relations différentes avec l'agent en question (flèches 4, 5 et 6). Ainsi, nous pouvons imaginer

ressembler ou être plus proches de l'agent en question. La force de ces trois relations permet cette fois la conversion de l'idée de P_2 en impression.

En conclusion, il semble que pour Hume le changement de point de vue nous permette de réviser nos jugements de deux manières : 1) il *peut* provoquer de nouvelles inférences concernant le caractère de l'agent (et, en conséquence, nous faire changer d'idée à ce sujet) ; 2) il modifie la force de la passion communiquée par la sympathie à cause du changement dans la force des relations qui nous lient au sujet du jugement.

2.3.4 Le contenu du jugement corrigé : les sentiments moraux

Comme nous avons vu dans la deuxième section de ce chapitre, une fois qu'on ressent la passion de l'agent par le biais de la sympathie limitée, cette impression provoque des sentiments d'approbation ou de désapprobation. Ces sentiments nous permettent de distinguer les vices des vertus. Mais Hume ajoute une précision à ce sujet :

Il est certain que ce sont nos *sentiments*, et non la *raison*, qui distinguent le bien et le mal en moral ; mais ces sentiments peuvent résulter soit [a] de la pure apparence ou du simple aspect des caractères et des passions, soit [b] de réflexions sur leur tendance à produire le bonheur de l'humanité et celui des personnes particulières (T 3.3.1.27, énumération ajoutée).

Dans ce passage, Hume décrit les deux circonstances qui peuvent donner lieu aux sentiments d'approbation ou de désapprobation. D'un côté, [a] lorsqu'on considère un caractère tout simplement, sans opérer un changement de point de vue. D'un autre côté, [b] lorsqu'on considère ce même caractère par le biais de la sympathie étendue, à savoir, en prenant en considération le point de vue d'autrui.

Dans ce deuxième cas, selon Hume, nous ressentons un type particulier de sentiments d'approbation ou de désapprobation, qu'il appelle *sentiments moraux*. Il affirme en effet que le sentiment moral « ne peut procéder que de notre sympathie avec les intérêts de la société » (T 3.3.1.12) et que :

Ce n'est que lorsqu'un caractère est considéré en général, sans référence à notre intérêt particulier, qu'il produit une impression ou un sentiment qui le désignent comme moralement bon ou mauvais (T 3.1.2.4).

Bref, ces sentiments diffèrent des sentiments d'approbation et de désapprobation concernant simplement un caractère, puisqu'ils découlent de la considération de ce caractère dans l'intérêt général de la société (T 3.3.1.30). Cette affirmation a comme implication que notre capacité à ressentir les sentiments moraux dépend de la vie en société. Cela dit, il ne faut pas oublier que par contre, selon Hume, notre capacité à ressentir de l'approbation ou de la désapprobation est originelle, à savoir elle est une propriété de la nature humaine (EPM 5.1.3-4).

En outre, Hume affirme que :

L'imagination s'attache aux vues *générales* des choses, et elle distingue les impressions [*feelings*] qu'elles produisent de celles qui résultent de notre situation particulière et momentanée (T 3.3.1.23).

Dans cet autre passage, Hume soutient que les sentiments qu'on ressent à partir de notre point de vue sont phénoménologiquement distincts de ceux qu'on ressent à partir du PVC. En d'autres mots, selon lui, nous sommes capables de distinguer les sentiments qui découlent d'une passion communiquée par la sympathie limitée, des sentiments provoqués par une passion qu'on ressent grâce à la sympathie étendue.

En conclusion, la sympathie étendue suscite des sentiments d’approbation et de désapprobation qui sont *les sentiments moraux*. Tout comme les sentiments d’approbation et de désapprobation, ces sentiments sont un type particulier de plaisir et de malaise sollicités de la considération d’un caractère.⁵⁴

2.3.5 L’idéal moral humien : le spectateur judicieux

J’ai commencé ce chapitre avec une citation tirée de l’*EPM*, où Hume affirme que pour ressentir les sentiments appropriés, nous devons bien raisonner. Cette idée n’est pas absente du *Traité*, où Hume dit que :

La nature humaine étant principalement composée de deux éléments qui sont nécessaires dans toutes ses actions – les passions et l’entendement –, il est certain que les mouvements aveugles des premières, s’ils ne sont dirigés par le second, rendent l’homme incapable de faire société (T 3.2.2.14).

Ce passage indique que la vie en société serait impossible si nos sentiments ne pouvaient pas être conduits par des raisonnements. Or, il est plausible de penser que les raisonnements en question sont ceux qui nous permettent de changer de perspective sur une situation et d’occuper le PVC comme j’ai précédemment illustré. En effet, lorsqu’on considère un caractère à partir du PVC, nous ressentons les sentiments moraux, et Hume affirme à ce propos que ces sentiments

⁵⁴ Pour une lecture différente au sujet de la nature des sentiments moraux voir Pall Árdal (2001). Dans cet article, Árdal soutient que les sentiments moraux ne sont pas un type particulier de plaisirs et de malaises, mais plutôt des passions indirectes (calmes) d’amour/haine/fierté/humilité. Contrairement à Árdal, je crois que les sentiments moraux sont toujours cooccurents aux passions indirectes de l’amour et de la haine, sans pourtant être identiques. Sur ce point je suis plutôt en accord avec Thomas Hearn (1973, pp. 290). Plus précisément, je crois que les passions indirectes, contrairement aux sentiments moraux, ne demandent pas nécessairement l’intervention du mécanisme sympathique. À ce propos, Loeb affirme que: « [...] sympathy need not have any role in the origin of indirect passions. (Note that sympathy is first introduced at II, i, 11, *after* the systematic discussion of the origin of pride and humility has been completed.) » (1977 pp. 401).

se fondent sur la considération de l'intérêt et de l'avantage de la vie en société (T 3.2.5.15).

Pour terminer, j'aimerais consacrer la dernière sous-section de mon mémoire à l'illustration de la manière dont le changement de point de vue peut être entendu comme étant automatique (dans cette sous-section j'utiliserai les termes automatique, inconscient et involontaire comme étant interchangeables). D'une part, pour la personne qui est inexpérimentée, le changement de point de vue est automatique dans la mesure où l'agent adopte le PVC suite à un désaccord avec autrui. L'adoption du PVC vise donc à mettre fin au malaise causé par le désaccord. D'autre part, pour la personne expérimentée, le changement de point de vue est automatique dans la mesure où l'agent se trouve à adopter le PVC afin de *prévenir* le malaise qui survient lorsque nous sommes dans le point de vue particulier. Conséquemment, l'adoption du PVC est automatique dans les deux cas, mais dans un cas elle advient de façon diachronique (c'est-à-dire après que nous ressentions le sentiment inapproprié et afin de le corriger), dans l'autre cas elle advient de façon synchronique.

Tout d'abord, comme nous avons vu précédemment, le changement de perspective est causé par un malaise dû à un conflit avec autrui. Donc cette opération de l'esprit ne nécessite pas l'intervention de la volonté. Comme je l'ai mentionné (voir note 17), Hume conçoit la volonté comme étant « l'impression interne que nous sentons et dont nous sommes conscients, lorsque nous suscitons sciemment un nouveau mouvement de notre corps ou une nouvelle perception de notre esprit » (T 2.3.1.2). De fait, il est possible d'imaginer qu'on arrive à occuper le PVC inconsciemment, simplement pour faire cesser la sensation déplaisante qui dérive du désaccord avec autrui. Cela dit, il faut souligner que d'aucune manière

Hume n'exclue que ce processus puisse aussi se produire consciemment. À cet effet, Herdt soutient que nous pouvons nous « efforcer à étendre notre sympathie au-delà de sa forme initiale limitée même si cette première sympathie est en fait involontaire » (1997, p. 74-75).

Ensuite, il y a plusieurs passages dans le *Traité*, aussi bien que dans l'*EPM* et l'*ENG*, où Hume dit qu'on peut aboutir à corriger les jugements en adoptant le PVC afin de prévenir les erreurs de jugement. Prenons un exemple afin de montrer comment nous maîtrisons cette méthode de correction des jugements selon Hume.

Supposons d'être au moment de la bataille des Thermopyles : l'armée grecque, conduite par le roi Léonidas, combat l'armée perse, guidée par Xerxès. Supposons que nous sommes des soldats perses présents à la bataille. Pendant le combat, une partie de l'armée grecque tue le roi Xerxès pour permettre aux autres la fuite, mais elle ne résiste pas à l'attaque perse qui s'ensuit. Il se peut donc que de notre point de vue particulier, les soldats grecs responsables d'avoir tué notre roi soient méprisables. Étant donné que ce jugement est formulé en partant de notre point de vue particulier, il sera erroné, selon notre interprétation. Hume affirme que :

Une cause évidente de ce que beaucoup ne parviennent pas à ressentir le véritable sentiment de la beauté est le manque de cette délicatesse d'imagination qui est requise pour prendre conscience de ces émotions fines (ENG 14).

Dans cet extrait, Hume soutient que, pour ressentir le *véritable* sentiment de la beauté il faut avoir une imagination raffinée. Or, pour Hume, il y a une certaine analogie dans le fonctionnement de notre sens esthétique et dans celui de notre sens moral (surtout dans l'*ENG*). Pour cette raison, nous pouvons supposer qu'il en va de même dans le cas moral : une des raisons pour lesquelles nous ne ressentons pas le sentiment moral est le manque de « délicatesse de

l'imagination ». Il faut maintenant comprendre comment nous pouvons acquérir cette *délicatesse de l'imagination* dont Hume parle dans cet extrait. Pour ce faire, continuons avec notre exemple et supposons que, dix ans après la bataille, nous soyons en train de discuter avec des vétérans grecs, d'après lesquels les soldats qui se sont sacrifiés lors de la bataille sont très honorables. Nous sommes évidemment en désaccord avec eux et cette sensation est déplaisante pour nous. Selon la théorie morale humienne, cela nous pousse à occuper le PVC qui nous permet de comprendre que, du point de vue des grecs, il est tout à fait normal de louer les soldats qui ont fait un tel sacrifice. C'est l'expérience du conflit qui « nous enseigne tôt cette méthode de correction de nos sentiments » dit Hume (T 3.3.1.16). Cela veut dire que la prochaine fois, nous occuperons le PVC non pas pour faire cesser la sensation déplaisante causée par le conflit avec autrui, mais plutôt pour l'éviter qu'elle puisse paraître en premier lieu. En outre, pour Hume, en corrigeant ainsi nos jugements, nous élargissons notre perspective sur les situations. Le plus de perspectives et de points de vue nous prenons en considération, le plus nous serons capables d'éviter les conflits avec autrui.

Enfin, pour Hume le spectateur judicieux peut aboutir à considérer un caractère dans toute sa complexité et dans ses détails en raisonnant à partir de plusieurs perspectives. En d'autres mots, en considérant le caractère de quelqu'un, le spectateur judicieux est capable d'en évaluer chaque aspect, les côtés vicieux et les côtés vertueux. Voici en quoi consiste la délicatesse de l'imagination à laquelle Hume faisait référence plus tôt. Hume l'affirme ainsi:

[...] plus nous nous habituons à un examen précis des questions morales, plus nous acquérons une perception [*feeling*] délicate des plus subtiles distinctions entre vice et vertu (EPM 5.1.14).

Et encore :

La grande ressemblance entre le gout de l'esprit et le gout physique nous apprendra aisément à tirer la leçon de cette histoire. Bien qu'il soit assuré que la beauté et la difformité, plus encore que le doux et l'amer, ne peuvent être des qualités inhérentes aux objets, mais sont entièrement le fait du sentiment interne ou externe, on doit reconnaître qu'il y a certaines qualités dans les objets qui sont adaptées par nature produire ces sentiments particuliers. Maintenant, comme ces qualités peuvent exister à un faible degré, ou bien peuvent être mélangées et confondues les unes avec les autres, il arrive souvent que le goût ne soit pas affecté par des traits aussi délicats, ou ne soit pas capable de distinguer toutes les saveurs particulières, dans le désordre où elles sont présentées. Là où les sens sont assez déliés pour que rien ne leur échappe, et en même temps assez aiguisés pour percevoir tout ingrédient introduit dans la composition : c'est là ce que nous appellerons délicatesse de goût, que nous employons ces termes selon leur sens littéral ou selon leur sens métaphorique (ENG 16).

2.4 Conclusion

Une partie de la littérature sur les biais cognitifs porte sur l'influence que le biais de l'affection a sur les jugements moraux. Dans ce mémoire, j'ai voulu proposer à ce sujet une réponse à la question suivante : est-il possible de corriger automatiquement les jugements qui sont influencés par le biais de l'affection ? Comme tentative de réponse à cette question, j'ai présenté dans le premier chapitre deux théories contemporaines concernant les mécanismes de correction des jugements moraux, à savoir celle de Greene et celle de Haidt. J'ai préféré la théorie de ce dernier, suivant laquelle nous pouvons corriger les jugements moraux par un changement de perspective sur la situation qu'on est en train d'évaluer moralement. J'ai pourtant remarqué que Haidt ne tient pas compte de la possibilité que ce mécanisme de correction des jugements moraux soit automatique. Dans ce deuxième chapitre, j'ai voulu montrer que, au contraire, Hume avait pris en considération une telle option. Il croit qu'il est possible qu'on apprenne à susciter automatiquement des jugements moraux qui reposent sur une bonne intuition, à savoir sur une intuition qui nous assure de vivre paisiblement en société.

En effet, dans ce chapitre j'ai défendu une lecture de la théorie morale humienne selon laquelle certains jugements moraux sont fondés sur des sentiments inappropriés, qu'il est préférable de corriger afin de vivre paisiblement en société. Ainsi, j'ai voulu mettre en évidence le fait que, selon Hume, nous pouvons corriger les erreurs de jugement à travers un changement de perspective qui devient automatique avec le temps et la pratique. D'ailleurs, la correction du jugement repose sur une correction du sentiment sur lequel le jugement repose. Afin d'illustrer ces aspects de la théorie morale humienne, j'ai articulé ce chapitre en trois sections.

Dans la première section, j'ai illustré le modèle humien du jugement moral qui se déroule en trois étapes. Nous y avons vu que la première étape dans la formulation d'une évaluation morale consiste à lire les intentions d'autrui, à savoir les passions qui le ou la motivent à l'action ; ensuite, la deuxième étape consiste à ressentir ces mêmes passions à travers la sympathie ; enfin, la troisième étape consiste à éprouver des sentiments d'approbation ou de désapprobation en réponse à la passion communiquée par le mécanisme sympathique.

Dans la deuxième section, j'ai montré que dans le cadre de la théorie humienne, nos évaluations morales peuvent être marquées par diverses erreurs de raisonnement.

Tout d'abord, nous pouvons nous tromper en lisant les intentions d'autrui, notamment lorsque nous raisonnons à partir de notre propre expérience sur le lien qu'il y a entre ses actions et ses motivations. Ce faisant on néglige de prendre en considération que les autres peuvent montrer leurs passions différemment de nous. Ensuite, produire une évaluation morale en nous basant seulement sur les passions passagères de celui ou celle que nous jugeons peut nous conduire en erreur. En

effet, selon Hume, nous devons plutôt prendre en considération les passions qui animent fréquemment autrui et non pas celles qui le motivent à l'action occasionnellement.

Finalement, j'ai voulu montrer que pour Hume nos erreurs de raisonnement dérivent du fait que nous considérons le caractère d'autrui à partir de notre propre point de vue, à savoir, par le biais de la *sympathie limitée*. J'ai en effet montré que pour Hume la sympathie limitée a comme point de départ notre propre point de vue, qui est intéressé et imbibé de préjugés. De ces préjugés peuvent découler des erreurs de raisonnement qui nous mènent à mésinterpréter les intentions d'autrui. De plus, la sympathie limitée nous porte à considérer la passion qui anime autrui à un seul instant, sans tenir en compte de la complexité de son caractère (Cunningham 2004, p. 239). Enfin, la sympathie limitée est biaisée par la force des relations qui nous lient à autrui, autrement dit nous sommes plus charitables à l'égard des actions des membres de notre famille qu'à l'égard des étrangers.

Dans un troisième moment, j'ai illustré comment et pourquoi, selon Hume, nous corrigeons ces erreurs de jugement.

J'ai initialement montré que la correction des jugements moraux advient à travers un changement de perspective qui a lieu inconsciemment (dans la plupart des cas) afin d'éviter le malaise qui découle d'un conflit d'opinions avec les autres ou d'une incohérence en nous-mêmes.

J'ai par ailleurs signalé que Hume ne spécifie pas clairement quel point de vue nous devons occuper afin d'être en accord les uns avec les autres. En effet, dans certains passages, il affirme qu'il faut adopter le point de vue de la personne qui est jugée, alors que dans d'autres il est plutôt question de son entourage, de la société ou bien encore de l'humanité toute entière.

Il me semble clair que pour Hume occuper le PVC signifie sympathiser avec l'agent par le biais de la *sympathie étendue*. J'ai donc argumenté afin de montrer

que sortir de son propre point de vue implique de remettre en question ses propres croyances concernant le lien entre les passions et les actions qui en dérivent.

D'ailleurs, la sympathie étendue nous pousse à prendre en considération le caractère d'autrui dans l'évaluation morale, nous faisant ainsi éviter les erreurs qui découlent d'une considération des passions passagères. En outre, la sympathie étendue nous porte à faire abstraction des relations personnelles qui nous lient aux autres en modifiant la force de ses dites relations.

Qui plus est, selon Hume, le changement de point de vue entraîne un changement dans nos sentiments qui nous porte à juger différemment une situation donnée. Dans ce cas, nous ressentons des sentiments moraux qui, pour Hume, sont les bons sentiments moraux à savoir, ceux qui nous permettent de communiquer avec autrui et de vivre en société.

Enfin, et surtout, selon Hume nous pouvons entraîner (délibérément ou inconsciemment) notre capacité à occuper le PVC, afin d'éviter les erreurs de jugement.

CONCLUSION

Ce mémoire a abordé deux questions étroitement liées. La première vise à déterminer la manière dont nous formulons nos jugements moraux, alors que la seconde a pour objectif de savoir comment nous pouvons nous assurer qu'ils ne sont pas biaisés par nos réponses affectives (subjectives).

En ce qui concerne la première question, dans le premier chapitre, j'ai pris en considération les deux modèles du jugement moral qui sont le plus influents dans la littérature contemporaine à ce sujet : celui de Greene et celui de Haidt. Dans chacun de ces deux modèles, il y a deux types de processus cognitifs qui interviennent dans la formulation de jugements moraux : les raisonnements (PT1) et les intuitions (PT2). Dans cette littérature, les intuitions sont le résultat d'heuristiques, c'est-à-dire des raccourcis que l'esprit emprunte afin de répondre à des stimuli de façon rapide et efficace. Bien qu'elles produisent souvent une réponse adéquate, les heuristiques peuvent pourtant nous induire en erreur. Lorsque c'est le cas, l'intuition qui en résulte sera biaisée, et il y a de forte chance qu'il en soit de même pour le jugement qui en découle.

En outre, nous avons vu qu'autant pour Greene que pour Haidt, les intuitions qui déterminent le contenu de nos jugements moraux ont un contenu affectif. Conséquemment, je considère plausible de soutenir que pour ces deux auteurs, les jugements moraux intuitifs sont le produit d'une heuristique de l'affect, qui peut parfois être biaisée.

Ensuite, en me penchant sur l'objectif explicatif de ces deux théories, j'ai tenté de mettre en relief leurs limites inhérentes. Dans le modèle de Greene, ce n'est que dans le cas où l'action qui est jugée risque de nuire à autrui que le contenu émotif des intuitions détermine nos jugements moraux (ce qu'il appelle, une « violation

morale personnelle »). Ce faisant, Greene ignore complètement d'autres types de situations qui peuvent provoquer de fortes réponses émotives et, ce faisant, possiblement déterminer le contenu de nos jugements.

Le modèle de Haidt, par contre, permet de prendre en compte de telles situations. Ainsi, ce modèle peut être considéré comme étant une expansion de celui de Greene, comportant deux différences principales à l'égard de ce dernier :

1. Pour Haidt, il y a plusieurs types situations qui peuvent provoquer une réponse émotive en nous, pas seulement des situations de violation morale personnelles ;
2. Haidt tient en compte de l'influence du contexte social dans la formulation des jugements moraux.

Le modèle de Haidt possède donc plus de portée explicative.

En ce qui concerne la deuxième question, celle de la révision des jugements moraux, j'ai illustré dans le premier chapitre comment, selon Greene et selon Haidt, nous pouvons corriger nos jugements moraux lorsqu'ils découlent du biais affectif. D'un côté, selon Greene, nous devrions douter de nos intuitions et raisonner de façon conséquentialiste sur l'action qui est objet de notre jugement. J'ai voulu souligner à ce propos que cette suggestion n'est praticable que dans l'ensemble limité de situations qu'il prend en considération dans son modèle, à savoir, les violations morales personnelles. De l'autre côté, selon Haidt, nous pouvons corriger notre jugement à l'aide d'un raisonnement logique ou bien d'un changement de perspective sur une situation donnée. Cette méthode peut être utilisée lors de situations qui ne sont pas des violations morales personnelles.

Par contre, étant donné que ces deux modèles reposent sur la théorie des processus-duaux de Kahneman, autant Greene que Haidt conçoivent que les processus nous permettant de corriger nos jugements moraux sont lents et demandent de l'effort afin d'être mis en place. Est-ce que cela veut dire que nous

n'apprendrons jamais à réviser automatiquement nos jugements moraux ? Haidt ne semble pas intéressé à développer davantage son explication des mécanismes et des stratégies de révision des jugements moraux. Pour cette raison, j'ai présenté la théorie humienne du jugement moral dans le deuxième chapitre de ce mémoire, selon laquelle nous pouvons réviser nos jugements moraux suite à un changement de point de vue que nous pouvons, avec la pratique, finir par adopter de façon synchronique à la formulation du jugement moral.

Dans le deuxième chapitre de ce mémoire, j'ai expliqué la théorie morale humienne en prêtant attention à trois aspects principaux.

En premier lieu, j'ai illustré le modèle du jugement moral conçu par Hume. Dans le cadre de la théorie morale humienne, les trois étapes dans la formulation du jugement moral consistent tout d'abord (1) à lire les passions qui ont motivé l'action d'autrui, ensuite (2) à ressentir ces mêmes passions par le biais de la sympathie et enfin (3) à éprouver des sentiments d'approbation ou de désapprobation dépendamment de la sensation que génère la passion et qui est transmise à la sympathie. Nous approuvons des passions qui nous causent un plaisir et nous désapprouvons des passions qui sont déplaisantes pour nous.

En deuxième lieu, j'ai voulu montrer que nous pouvons commettre trois types d'erreurs de jugements selon Hume. En outre, j'ai montré qu'ils découlent tous les trois du fait que nous jugeons quelqu'un à partir de notre propre point de vue. Ainsi, nous pouvons nous tromper en lisant les intentions d'autrui à partir des préjugés qui caractérisent notre perspective particulière. Nous pouvons également nous tromper en prenant en considération des passions passagères chez autrui plutôt que son caractère au complet. Cette erreur est plus probable si notre jugement dérive d'une sympathie limitée puisque cette forme de sympathie implique que nous recevons par communication la passion qui anime autrui dans un seul instant. Enfin, la dernière erreur découle du fait que la sympathie limitée

fait en sorte que nous sommes plus cléments dans nos jugements qui portent sur ceux qui nous sont proches qu'avec ceux qui sont loin de nous.

En troisième lieu, j'ai illustré comment, selon Hume, nous pouvons réviser nos jugements moraux. Nous avons vu que la correction des jugements moraux dépend de notre capacité à transcender notre point de vue particulier afin d'occuper le PVC. Étant donné que la sympathie étendue implique que nous occupions le point de vue d'autrui plutôt que le nôtre il en suit que la sympathie étendue implique que nous occupions le PVC. Ceci rappelle le type de correction des jugements moraux conçu par Haidt, selon qui nous pouvons réviser nos jugements moraux en changeant de point de vue sur la situation donnée.

De plus, j'ai soutenu que, conformément à la théorie humienne, lorsque nous adoptons le PVC pour juger quelqu'un, nous sommes poussés à mettre en question nos conclusions concernant les intentions de l'agent en question. En effet, en adoptant le PVC nous transcendons notre point de vue particulier imbibé de préjugés. En outre, la sympathie étendue nous pousse à prendre en considération le caractère des sujets avec qui nous sympathisons plutôt que leurs passions passagères. Enfin, cette forme de sympathie n'est pas biaisée par nos relations personnelles avec l'agent. De cette manière, la sympathie étendue nous permet de réviser nos jugements moraux.

Ensuite, nous avons vu que l'adoption du PVC est motivée par notre aversion au malaise provoqué par le désaccord avec les autres membres de notre société. Cela veut dire que, pour Hume, l'intervention de la volonté n'est pas nécessaire pour que nous occupions ce PVC. Enfin, il croit que la pratique et l'exercice de notre capacité à approuver ou à désapprouver un caractère nous conduit à changer de point de vue sur la situation de façon automatique, évitant ainsi les erreurs de jugement du départ. Cela nous permettra de communiquer et de vivre paisiblement en société selon Hume, qui semble ainsi soutenir le même point de vue que Greene, ainsi que de Dunbar, à propos de la valeur des jugements moraux dans la constitution de notre société.

Ce mémoire avait comme but de suggérer une nouvelle méthode afin de prévenir les erreurs dans les jugements moraux (notamment ceux causés par le biais de l'affect). Cette méthode repose sur la suggestion humienne qui consiste à entraîner notre capacité à sortir de notre point de vue particulier. Je crois qu'il est au moins plausible de croire que le processus psychologique nous permettant de quitter notre point de vue et d'occuper le point de vue d'autrui ressemble à la prise de rôle qui, selon Haidt, nous permet de corriger nos jugements moraux intuitifs. Plus précisément, pour Hume comme pour Haidt ce changement de perspective provoque une nouvelle réaction affective. Je crois conséquemment que le prochain pas à faire serait celui d'explorer les recherches empiriques (en psychologie et en neurosciences) afin de comprendre si nous pouvons effectivement entraîner notre capacité à occuper le point de vue d'autrui et si cette habitude nous permettrait de corriger préventivement nos jugements moraux lorsqu'ils sont biaisés.

BIBLIOGRAPHIE

Árdal, P.S. 2001. « Another Look at Hume's Account of Moral Evaluation ». *The Journal of the History of Philosophy*, Vol. 15, No. 4, Octobre, pp. 405-421.

Baier, A.C. 1991. *A Progress of Sentiments*, Harvard University Press. 333p.

Chap, J. B. 1986. « Moral Judgment in Middle and Late Adulthood: The Effects of Age-Appropriate Moral Dilemmas and Spontaneous Role Taking ». *The International Journal of Aging and Human Development*, Vol. 22, No. 3, pp. 161-172.

Cohon, R. 1997. « The Common Point of View in Hume's Ethics ». *Philosophy and Phenomenological Research*, Vol. 57, No. 4, Décembre, pp. 827-850.

Cunningham, A.S. 2004. « The Strenght of Hume's "Weak Sympathy" ». *Hume Studies*, Vol. 30, No. 2, pp. 237-256.

D'arms J., Jacobson, D. 2000. « The Moralistic Fallacy: On the 'Appropriateness' of Emotions ». *Philosophy and Phenomenological Research*, Vol. 61, No. 1, pp. 65-90.

Debes, R. 2007a. « Humanity, Sympathy and the Puzzle of Hume's Second Enquiry ». *British Journal for the History of Philosophy*, Vol. 1, No. 1, pp. 27-57.

Debes, R. 2007b. « Has Anything Changed? Hume's Theory of Association and Sympathy after the Treatise ». *British Journal for the History of Philosophy*, Vol. 15, No. 2, pp. 313-338.

Debes, R. 2012. « Recasting Scottish Sentimentalism: The Peculiarity of Moral Approval ». *Journal of Scottish Philosophy*, Vol. 10, No. 1, pp. 91-115.

Dunbar, R. I. M. 2004. « Gossip in Evolutionary Perspective ». *Review of General Psychology*, Vol. 8, No. 2, pp. 100-110.

Eisenberg, N. Zhou, Q. Koller, S. 2001. « Brazilian Adolescents' Prosocial Moral Judgment and Behavior: Relations to Sympathy, Perspective Taking, Gender-Role Orientation, and Demographic Characteristics ». *Child Development*, Vol. 72, No. 2, pp. 518-534.

Eskine, K.J. Kacinik, N. A et Prinz, J.J. 2011. « A Bad Taste in the Mouth: Gustatory Disgust Influences Moral Judgment ». *Psychological Science*, Vol. 22, No. 3, pp. 295-299.

Evans, J. 2008. « Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition ». *Annual Review of Psychology*, Vol. 59, pp. 255-278.

Evans, J. and Frankish, K. 2009. « The Duality of Mind : An Historical Perspective ». Dans J. Evans et K. Frankish (dir.), *Two Minds: Dual Processes and Beyond*, Oxford: Oxford University Press, pp.1-29.

Evans, J. 2012. « Spot the Difference: Distinguishing Between two Kinds of Processing ». *Mind & Society*, Vol. 11, pp. 121-131.

Faucher, L. Tappolet, C. 2008. « Facts and Values in Emotional Plasticity ». Dans L. C. Charland et P. Zachar (dir.), *Fact and Value in Emotion*. pp. 101-137.

Frankish, K. 2010. « Dual-Process and Dual-System Theories of Reasoning ». *Philosophy Compass*, Vol. 5, pp. 914-926.

Gawronski, B. Sherman, J. W. Trope, Y. 2014. « Two of What? A Conceptual Analysis of Dual-Process Theories ». Dans J. W. Sherman, B. Gawronski et Y. Trope (dir.), *Dual-process theories of the social mind*, pp. 3-19.

Gigerenzer, G. 2008. *Rationality for Mortals: How People Cope with Uncertainty*. Oxford: Oxford University Press.

Gilovich, T., Griffin, D. Kahneman, D. 2002. *Heuristic and Biases: The Psychology of Intuitive Judgment*.

Greene, J.D. 2009. « Dual-Process Morality and the Personal/Impersonal Distinction: A Reply to McGuire, Langdon, Coltheart, and Mackenzie ». *Journal of Experimental Social Psychology*, Vol. 45, No. 3, pp. 581-584.

Greene, J.D. 2013. *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. The Penguin Press. New York.

Greene, J.D. 2014. « Beyond Point-and-Shoot Morality : Why Cognitive (Neuro)Science Matters for Ethics ». *Ethics*. Vol. 124, No. 4, pp. 695-726.

Greene, J.D. et al. 2001. « An fMRI Investigation of Emotional Engagement in Moral Judgment ». *Science*, Vol. 293, pp. 2105-2108.

Greene, J.D. & Haidt, J. 2002. « How (and Where) Does Moral Judgment Work? ». *Trends in Cognitive Sciences*, Vol. 6, No. 12, pp. 517-523.

Greene, J.D. Nystrom, L.E. Engell, A.D. Darley, J.M. et Cohen, J. D. 2004. « The Neural Bases of Cognitive Conflict and Control in Moral Judgment ». *Neuron*, Vol. 44, pp. 389–400.

Haidt, J. 2001. « The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment ». *Psychological Review*. Vol. 108, pp. 814-834.

Haidt, J. 2003. « The Moral Emotions ». Dans R. J. Davidson, K. R. Sherer et H. H. Goldsmith (dir.), *Handbook of Affective Sciences*, pp. 852-870.

Haidt, J. 2006. *The Happiness Hypothesis : Finding Modern Truth in Ancient Wisdom*. New York : Basic Books.

Haidt, J. 2007. « The New Synthesis in Moral Psychology ». *Science*, Vol. 316, pp. 998-1002.

Haidt, J. 2012a. « Moral Psychology and the Law : How Intuitions Drive Reasoning, Judgment, and the Search for Evidence ». *Alabama Law Review*, Vol. 64, No. 4, pp. 867-880.

Haidt, J. 2012b. *The Righteous Mind : Why Good People Are Divided by Politics and Religion*. New York : Pantheon Books.

Haidt, J. 2013. « Moral Psychology for the Twenty-first Century ». *Journal of Moral Education*, Vol. 42, No. 3, pp. 281-297.

Haidt, J. Koller, S. & Dias, M. 1993. « Affect, Culture, and Morality, or is it Wrong to Eat Your Dog? ». *Journal of Personality and Social Psychology*, Vol. 65, No. 4, octobre, pp. 613-628.

Hausman, D. 1988. « Can Hume's Use of a Simple/Complex Distinction Be Made Consistent? ». *Hume Studies*, Vol. 14, No. 2 (Novembre), pp. 424- 428.

Herd, J.A. 1997. *Religion and Faction in Hume's Moral Philosophy*. Cambridge University Press. 300 p.

Huebner, B. Dwyer, S. et Hauser, M. 2008. « The Role of Emotion in Moral Psychology ». *Trends in Cognitive Sciences*, Vol. 13, No. 1, pp. 1-6.

Hume, David. 1998. *An Enquiry Concerning Human Understanding*, Tom L. Beauchamp, The Clarendon Edition of the Works of David Hume, Oxford, Oxford University Press.

Hume, David. 1999. *An Enquiry Concerning the Principles of Morals*, Tom L. Beauchamp, Oxford, Oxford University Press.

Hume, David. 2007. *A Treatise of Human Nature*, David Fate Norton et Mary J. Norton, Oxford, Clarendon Press.

Hume, David. 2006. *Enquête sur l'Entendement Humain*, traduction de A. Leroy et M. Beyssade éditée par Flammarion.

Hume, David. 1991. *Enquête sur les Principes de la Morale*, traduction de P. Baranger et P. Saltel, éditée par Flammarion.

Hume, David. 2000. *Essai sur la Norme du Goût*, traduction de R. Bouveresse éditée par Flammarion.

Hume, David. 1985. *Of The Standard of Taste*. Eugene F. Miller, Indianapolis: Liberty Classics.

Hume, David. 1995. *Traité de la Nature Humaine*, Livre I, trad. Ph. Baranger et Ph. Saltel, Flammarion, Paris.

Hume, David. 1991. *Traité de la Nature Humaine*, Livre II, trad. Ph. Baranger et Ph. Saltel, Flammarion, Paris.

Hume, David. 1993. *Traité de la Nature Humaine*, Livre III, trad. Ph. Baranger et Ph. Saltel, Flammarion, Paris.

Kahneman, D. et Tversky, A. 1974. « Judgment under Uncertainty: Heuristics and Biases ». *Science*, New Series, Vol. 185, No. 4157, p. 1124-1131.

Kahneman, D. Frederick, S. 2002. « Representativeness Revisited: Attribute Substitution in Intuitive Judgment ». Dans T. Gilovich, D. Griffin et D. Kahneman (dir.) *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press. pp. 49–81.

Kahneman, D. 2003. « Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice ». *The American Economic Review*, Vol. 93, No. 5, pp. 1449-1475.

Kelly, D. Stich, S. Haley, K.J. Eng, S.J. & Fessler, D.M.T. 2007. « Harm, Affect, and the Moral/Conventional Distinction ». *Mind & Language*, Vol. 22, No. 2, pp. 117-131.

Kenyon, T. et Beaulac, G. 2014. « Critical Thinking Education and Debiasing ». *Informal Logic*. Vol. 34, No. 4, pp. 341-363.

Koenings, M. Young, L. Adolphs, R. Tranel, D. Cushman, F. Hauser, M. Damasio, A. 2007. « Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments ». *Nature*, Vol. 446, No. 7138, pp. 908–911.

Korsgaard, C.M. 1999. « The General Point of View: Love and Moral Approval in Hume's Ethics ». *Hume Studies*, Vol. 25, No. 1 et 2, Avril/Novembre, pp. 3-41.

Landy, D. 2006. « Hume's Impression/Idea Distinction ». *Hume Studies*, Vol. 32, No. 1, pp. 119-140.

Maibom, H. 2010. « What Experimental Evidence Shows Us about the Role of Emotions in Moral Judgement ». *Philosophy Compass*, Vol. 5, No. 11, pp. 999–1012.

Nichols S. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*, Oxford University Press S - New York.

Norton, D.F. 1984. « Hume's Moral Ontology ». *Hume Studies*, 10th Anniversary Issue, pp. 189-214.

Paxton, J.M. Ungar, L. Greene, J.D. 2012. « Reflection and Reasoning in Moral Judgment ». *Cognitive Science*, Vol. 36, pp. 163-177.

Pitson, T. 1989. « Projectionism, Realism, and Hume's Moral Sense Theory ». *Hume Studies*, Vol. 15, No. 1, Avril, pp. 61-92.

Pitson, T. 2002. « Hume on Character and the Self ». Dans T. Pitson, *Hume's Philosophy of the Self*, Routledge: Londre, pp. 83-101.

Prinz, J.J. 2006. « The Emotional Basis of Moral Judgments ». *Philosophical Explorations*, Vol. 9, No. 1, pp. 29-43.

Prinz, J.J. 2009. « Emotions: Motivating Feelings ». Dans B. McLaughlin, A. Beckermann et S. Walter (dir.), *The Oxford Handbook of Philosophy of Mind*, pp. 678-690.

Remow, G. 2009. *Hume's Moral Realism*. Dissertation/thèse, Université de Washington.

Samuels, R. 2009. « The Magical Number Two, Plus or Minus: Comments on Dual Systems ». Dans J. Evans, et K. Frankish (dir.), *In Two Minds: Dual Processes and Beyond*, OUP, pp. 129-149.

Saunders L. F. 2009. « Reason and Intuition in the Moral Life: A Dual Process Account of Moral Justification ». Dans J. Evans et K. Frankish (dir.) *In Two Minds: Dual Processes and Beyond*, OUP, pp. 335-355.

Schmicking, D.A. 2004. « Hume's Theory of Simple Perceptions Reconsidered ». *Hume Studies*, Vol. 30, No. 1, pp. 3-31.

Seidel, A. Prinz, J.J. 2013. « Mad and Glad: Musically Induced Emotions Have Divergent Impact on Morals ». *Motivation and Emotion*, Vol. 37, pp. 629-637.

Slovic, P. Finucane, M.L. Peters, E. MacGregor, D.G. 2007. « The Affect Heuristic ». *European Journal of Operational Research*, Vol. 177, pp. 1333–1352.

Strohming, N. Lewis, R.L. Meyer, D.E. 2011. « Divergent Effects of Different Positive Emotions on Moral Judgment ». *Cognition*, Vol. 119, pp. 295-300.

Sustein, C.R. 2005. « Moral Heuristics ». *Behavioral and Brain Sciences*, Vol. 28, pp. 531-573.

Wheatley, T., & Haidt, J. 2005. « Hypnotically Induced Disgust Makes Moral Judgments More Severe ». *Psychological Science*, Vol. 16, pp. 780-784.

Wilson, T.D. Brekke, N. 1994. « Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations ». *Psychological Bulletin*, Vol. 116, No. 1, pp. 117-142.