

## **Jugements moraux et motivation à la lumière des données empiriques**

CHRISTINE CLAVIEN  
UNIVERSITE DE LAUSANNE  
christine.clavien@unil.ch

### **Résumé**

Cet article contient un « tableau affectif », une analyse détaillée, largement basée sur des données empiriques, de la manière dont les gens jugent et se comportent moralement, où sont précisés les rôles respectifs des processus réflexifs et affectifs dans l'activité morale. Il est utile de distinguer différentes sortes de jugements, certains étant plus complexes que d'autres. Les « jugements sophistiqués » interviennent dans le cadre d'une activité rationnelle alors que les « jugements de valeur spontanés » sont une manière primitive et non réflexive d'évaluer le monde. Cette analyse permettra de prendre position dans le traditionnel débat sur l'internalisme et l'externalisme. Je défendrai l'idée que les normes et les jugements moraux que nous défendons n'exercent pas d'effet particulier sur nos actions si nous n'avons pas, par ailleurs, une attitude émotionnelle correspondante. Or cette dernière semble relever davantage de nos jugements spontanés.

*Mots clef : affect, émotion, réaction émotionnelle, externalisme, internalisme, justification, jugement moral, motivation, norme, valeur*

---

### **1. Le problème de la motivation**

Raymond croit en l'obligation morale d'aider toute personne en détresse. En se promenant un dimanche matin, il voit un homme étendu de tout son long sur le trottoir. Il est clair que l'homme a été battu et qu'il souffre horriblement. Alors que Raymond aurait le temps et les moyens de se porter en aide, il décide sans raison particulière de passer son chemin. Ce faisant, il est parfaitement conscient du fait qu'il transgresse une

norme morale à laquelle il souscrit ; en d'autres termes, il agit consciemment à l'encontre de ce qu'il pense devoir faire.

Cet exemple est assez troublant. Quoique parfaitement crédible, il contredit l'idée souvent défendue que le fait d'accepter une norme et les jugements qui en découlent, est un facteur motivant à l'action. C'est du moins l'avis des défenseurs de la position internaliste selon laquelle les considérations morales – en particulier le fait de produire un jugement moral – motivent les sujets à agir conformément à ces considérations. Un internaliste<sup>1</sup> défend l'idée qu'il existe une connexion interne ou nécessaire entre nos jugements moraux et la motivation à produire les actions prescrites par ces jugements. Cette position est compatible, soit avec le point de vue selon lequel les jugements moraux motivent par eux-mêmes, ou avec le point de vue selon lequel ils sont nécessairement mais indirectement connectés à la motivation.

Le comportement de Raymond entre en contradiction directe avec l'internalisme et semble confirmer la position externaliste selon laquelle, même si l'on peut établir une relation entre le jugement et la motivation, il s'agit d'une connexion externe et contingente ; en d'autres termes, les jugements moraux ne possèdent aucune force motivante en eux-mêmes pas plus qu'ils ne sont *nécessairement* connectés à la motivation.

Un internaliste pourrait objecter que l'exemple de Raymond n'est pas réaliste ; ou alors, il pourrait être tenté d'ajouter des détails à l'histoire – il pourrait ajouter par exemple que Raymond était en réalité incliné à aider l'homme blessé mais d'autres

---

<sup>1</sup> Tel qu'il est décrit ici, l'internalisme s'approche des positions auxquelles s'intéressaient David Brink et Jonathan Dancy. “[internalism states that] moral considerations necessarily motivate” (Brink 1989, p. 42). “[It is] the conception according to which it is impossible for an agent to make a sincere cognitive moral judgement and not to be motivated accordingly” (Dancy 2000, p.21).

Certains auteurs ont raffiné la thèse internaliste en y intégrant la condition supplémentaire de la rationalité pratique : la connexion internaliste est supposée fonctionner uniquement pour des agents doués de raison pratique (Smith 1996). Cette position ne sera pas discutée dans cet article parce qu'elle s'éloigne à mon avis de la question originale, d'autant plus qu'il n'est pas évident de saisir ce qu'il faut entendre par « agent doué de raison pratique ». En revanche il vaut la peine d'ajouter une clause d'exception en faveur de l'internalisme qui exclut de la discussion toute situation où l'agent souffre de désordres motivationnels susceptibles de l'affecter durablement – par exemple la dépression (Svavarsdottir 1999).

motifs égoïstes l'en ont empêché. Ces tentatives de décrédibiliser ou de redécrire les exemples dérangeants sont évidemment rejetées par les externalistes ; selon eux, de telles pratiques reviennent à un pur et simple refus de traiter la question.

Le débat entre l'internalisme et l'externalisme a une longue histoire et ce n'est pas mon intention de répéter ici des arguments déjà connus. Mon propos consistera à soutenir l'idée qu'il est possible de prendre au sérieux des exemples comme celui de Raymond tout en résolvant le débat sur de la motivation morale au moyen d'une simple distinction entre deux sortes de jugements. Je tâcherai de montrer que l'internalisme est vrai pour les « jugements spontanés » et l'externalisme pour les « jugements sophistiqués ». Plus précisément, je soutiendrai que la motivation ne provient pas de considérations d'ordre conceptuel – ce qui est le cas des jugements moraux sophistiqués – mais plutôt d'évaluations rapides et non réflexives, lesquelles sont directement responsables de nos jugements évaluatifs spontanés. Cette thèse sera défendue en deux étapes.

Il s'agira d'abord de présenter et développer un « tableau affectif » qui décrit la manière dont les gens évaluent des situations et forment leurs propres valeurs et normes. Sans chercher de preuves philosophiques pour soutenir les détails de ce tableau, je tâcherai d'en consolider les traits en recourant à des données et explications empiriques issues de diverses sciences comme la psychologie, l'anthropologie, l'économie empirique ou les neurosciences.

Dans un second temps, je montrerai que mon tableau affectif pose un nouveau cadre pour réfléchir à la question de la motivation ; nous verrons que lorsque qu'il est combiné à une hypothèse humienne sur la motivation, ce cadre permet de remettre sérieusement en question la position internaliste et, dans une moindre mesure, l'externalisme. Cela nous permettra de comprendre pourquoi la décision de Raymond est à la fois possible et déroutante.

## **2. Les grandes lignes du tableau affectif**

Afin de résoudre le cas de Raymond, je vais élaborer un *tableau affectif* qui consiste à mettre en évidence les ingrédients et mécanismes psychologiques sous-

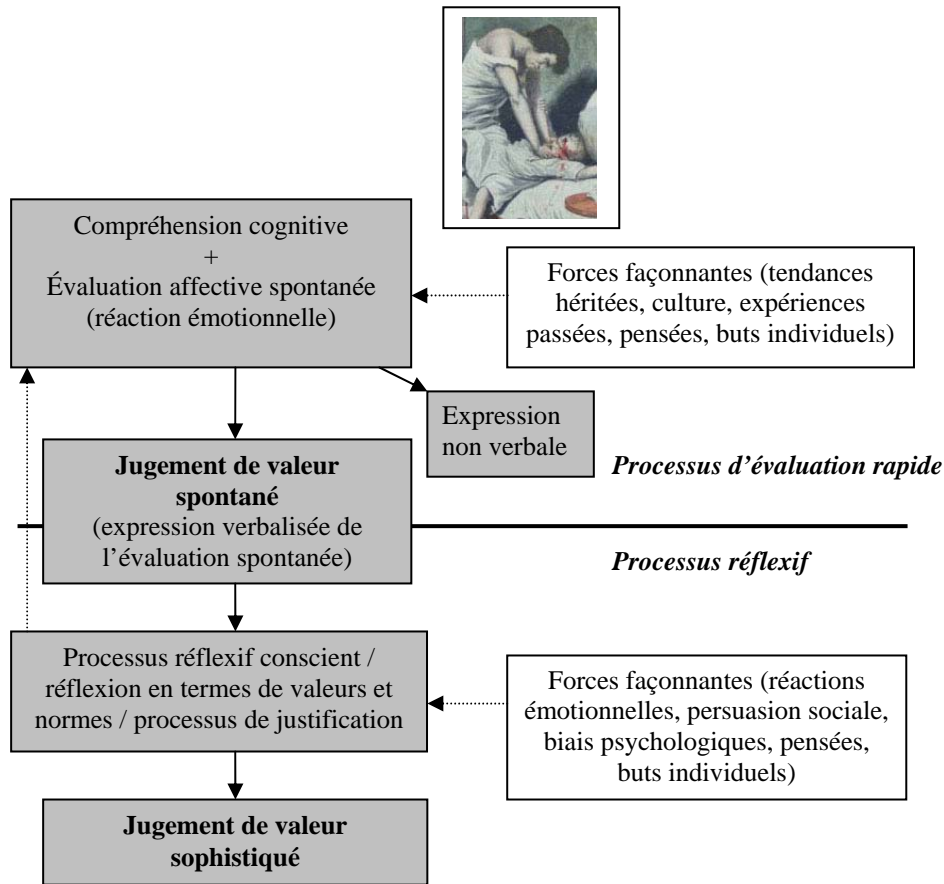
jacents à l'activité morale.<sup>2</sup> Il s'agira de proposer dans un premier temps, une description de ce qui se passe dans nos esprits lorsque nous jugeons une situation comme devant être réalisée ou évitée, et dans un second temps, de déceler ce qui nous motive à agir. La notion de « tableau affectif » réfère au fait que l'analyse se situe au niveau purement *descriptif* – il s'agit de dépeindre de manière systématique la pensée et l'activité morales – et accorde une place particulière à l'aspect *affectif*.

Le tableau affectif n'entre pas sous la coupe de deux conceptions classiques qui accordent une large place à l'affect : l'émotivisme pour qui les jugements moraux sont des expressions de réactions émotionnelles (Ayer 1946/1936), et l'expressivisme qui considère les jugements moraux comme des expressions d'états d'esprit au sujet d'émotions (Gibbard 1990).<sup>3</sup> Il me semble que ces deux manières de concevoir les jugements moraux minimisent excessivement notre capacité de mener des raisonnements complexes et, par le moyen de processus d'inférence et *sans* l'aide des émotions, de tirer des conclusions au sujet de ce qu'il faut faire. Nous verrons que l'analyse du jugement moral proposée ici est plus complexe. Voici un schéma qui sera explicité dans les sections qui suivent.

---

<sup>2</sup> Pour les besoins de cet article, je ferai usage d'une conception large de la notion de moralité, qui comprend toutes les réactions, réflexions et décisions relatives aux devoirs liés au bien-être et aux besoins des êtres humains. De même, la notion de jugement sera utilisée en un sens large : une simple expression de réaction émotionnelle peut être considérée comme un jugement. Ce choix n'est pas problématique dans la mesure où il existe une abondante littérature en philosophie morale qui conçoit les jugements moraux de cette manière – par exemple le courant émotiviste.

<sup>3</sup> Selon la théorie de Gibbard (1990) par exemple, les jugements moraux sont des expressions d'états d'esprit complexes qui consistent dans le fait de penser qu'il est justifié de ressentir une certaine émotion face à un certain type de situation.



Les flèches représentent des influences causales, soit directes (flèches pleines), soit diffuses (flèches en pointillé). Les boîtes grises représentent des activités de l’agent, et les boîtes blanches représentent les influences extérieures.

En deux mots, les grandes lignes de mon tableau affectif sont les suivantes : il y a deux processus dans l’activité morale – ou plus généralement dans l’activité évaluative. Le premier, que j’appellerai « processus d’évaluation rapide », est émotionnel et automatique alors que le second, que j’appellerai « processus réflexif », plus raffiné, inclut les réflexions conscientes et raisonnements d’inférence. Dans cet article, j’avancerai l’idée que seul le premier processus est capable de nous motiver à agir alors que seuls les éléments du processus réflexif peuvent être justifiés. En présentant ce double modèle, je montrerai qu’il existe différents types de jugements et que les plus sophistiqués n’ont pas le pouvoir de motiver à l’action.

### 3. Données empiriques sur les jugements moraux

Cette section est consacrée à la présentation d'une collection de données issues de la psychologie et des neurosciences, y compris certaines expériences en imagerie cérébrale. Ces données contribuent à éclairer le schéma présenté ci-dessus et les relations qu'il établit entre nos réactions émotionnelles et nos choix réflexifs sophistiqués. Nous verrons que les études empiriques nous aident à comprendre la relation entre les jugements moraux et les émotions. Elles montrent notamment que les jugements moraux ne résultent généralement pas de processus d'inférence au cours desquels nous appliquons consciemment une norme ou une valeur à une situation ; il semblerait plutôt que la plupart des évaluations soient intuitives, émotionnelles, largement automatisées.

Pour commencer, un ensemble d'études brise le vieux mythe de la rationalité de nos esprits moraux et soutient la distinction proposée entre le processus d'évaluation rapide et le processus réflexif.

L'étude la plus connue a été faite par Jonathan Haidt (2001) et montre que la plupart des gens condamnent de manière impulsive les pratiques incestueuses ou d'autres violations de tabous et soutiennent ce verdict même au terme d'une discussion où ils sont forcés d'admettre qu'ils ne disposent d'aucune bonne raison pour fonder leur jugement. Selon Haidt, ces résultats mettent en doute la rationalité de la moralité en montrant qu'une part non négligeable de notre activité morale ne peut pas être influencée par un processus réflexif. En d'autres termes, un groupe important de jugements moraux semble se fonder sur de simples réactions intuitives et non sur des raisons morales.

Il me semble que la manière la plus éclairante d'interpréter ces résultats est de comprendre nos évaluations intuitives comme des réactions émotionnelles ; on pourrait les comprendre comme des évaluations affectives spontanées. Nous verrons à la section suivante que cette idée de faire des émotions un constituant essentiel à l'activité morale n'est pas neuve. Mais pour le moment, considérons d'autres données empiriques susceptibles de nous apporter des informations intéressantes au sujet de la manière intuitive et émotionnelle dont nous évaluons les états de faits.

Thalia Wheatley et Jonathan Haidt (2005) ont mené une expérience sur des sujets hautement hypnotisables. Sous condition d'hypnose, ils ont expliqué aux sujets qu'ils éprouveraient du dégoût chaque fois qu'ils liraient un mot arbitraire – par exemple le mot 'donc'. Une fois réveillés, ils ont demandé aux sujets de lire et juger moralement une série de petites histoires assez communes – dont certaines n'étaient même pas particulièrement pertinentes au plan moral – qui contenaient ou non le mot lié au dégoût hypnotique. Les résultats sont impressionnants : systématiquement, les sujets condamnent moralement les histoires contenant les mots qui causent en eux du dégoût. Cette étude montre que des émotions causées de manière artificielle modulent nos jugements et parfois même génèrent des réactions morales.

Dans une expérience congruente menée par Greene et collègues (2001), les sujets doivent mener des expériences de pensée sur différentes variantes de situations problématiques comme celle du trolleybus où il s'agit de choisir de tuer une personne afin d'éviter la mort de cinq autres personnes.<sup>4</sup> Au cours de l'expérience, on demande aux sujets quelle action, parmi deux options données, ils considèrent comme moralement adéquate. Le cerveau des sujets est scanné à l'aide de la technique d'imagerie cérébrale afin de détecter les zones du cerveau qui s'activent. Les résultats montrent que l'engagement émotionnel – révélé par une activité accrue des zones du cerveau liées aux émotions – influence largement les jugements moraux : s'imaginer devoir pousser une personne sous un trolleybus roulant à pleine vitesse afin de le stopper, et par là, sauver cinq passagers, est émotionnellement plus saillant que s'imaginer pousser une manette qui va orienter la trajectoire du trolleybus sur une voie où se trouve un homme – dans cette deuxième option, la vie de l'homme sera également sacrifiée pour stopper le trolleybus. Cette différence d'engagement émotionnel induit les sujets à condamner la première action et à juger la seconde comme moralement permissible, alors même que dans les deux cas, la vie d'une personne est sacrifiée pour sauver celle de cinq autres.

La manière la plus simple d'interpréter ces données est de dire que l'engagement émotionnel influence fortement nos jugements moraux. Toutefois, il est important de remarquer que ces données ne permettent pas de prouver que les réactions

---

<sup>4</sup> Le premier modèle de cette expérience de pensée est dû à la philosophe Philippa Foot.

émotionnelles exercent un effet causal sur les jugements moraux ; on pourrait imaginer qu'elles soient des effets secondaires, eux-mêmes causés par les jugements. Cette position est défendue par une série de chercheurs qui ont mené des études similaires sur le dilemme du trolleybus ou des dilemmes similaires. Voici un résumé de leurs résultats et de la manière dont ils les interprètent.

Dans différentes études, des psychologues ont testé la manière dont les gens justifient leurs jugements moraux (Mikail 2002 ; Hauser 2006 ; Cushman *et al.* 2006). On présente aux sujets des problèmes moraux comme celui du trolleybus. Puis, on leur demande dans un premier temps quelles actions ils considèrent comme moralement acceptables ou moralement condamnables et dans un deuxième temps comment ils justifient leur jugement. L'analyse des réponses données par les sujets suggère que certaines règles intuitives profondément ancrées dans nos esprits guident les jugements des gens même si elles ne font pas surface dans leurs raisonnements conscients ; lorsqu'il s'agit de justifier leurs évaluations spontanées, les sujets n'invoquent pas toujours les règles qui les semblent les avoir poussés à formuler ces jugements. Parmi ces règles intuitives et parfois même inconscientes, il y a notamment celle selon laquelle un tort causé de manière intentionnelle en tant que moyen pour atteindre une fin est moralement plus condamnable que le même tort s'il est conçu comme l'effet secondaire d'un but que l'on cherche à atteindre.<sup>5</sup> Comme autre exemple, les gens conçoivent un tort causé directement avec contact physique comme moralement plus condamnable que le même tort causé sans contact physique.<sup>6</sup>

Des auteurs comme John Mikail pensent que ces résultats sont incompatibles avec l'idée que l'engagement émotionnel influence le jugement moral. Marc Hauser considère que ces données sur les jugements intuitifs supportent la position selon laquelle les émotions ne précèdent pas mais découlent des nos jugements moraux. Selon lui, un système gouverné par des règles déclenche une évaluation et cette évaluation déclenche parfois – mais pas systématiquement – une réponse émotionnelle. Ce système gouverné par des règles serait un sens du bien et du mal qui a évolué au cours de

---

<sup>5</sup> Il s'agit en fait de la doctrine du double effet qui avait déjà été introduite par Thomas d'Aquin 1985/1265-1273, Qu. 64, Art.7

<sup>6</sup> Pour des résultats similaires, voir aussi les travaux de Baron 1998 et de Nisbett & Wilson 1977.

millions d'années et serait partagé par l'ensemble des êtres humains ; il précède à la fois les jugements et les émotions.

Toutefois, il convient de noter que les données présentées par ces auteurs ne fournissent pas des informations permettant de *prouver* une telle interprétation. Il me semble que ce débat repose sur une mécompréhension. En observant de plus près les données et les interprétations présentées dans les deux groupes de recherche, on constate qu'elles sont parfaitement compatibles. Pour peu que l'on considère les réactions émotionnelles comme des *actes d'évaluation*, les deux interprétations peuvent être combinées. Les résultats de toutes les expériences sont parfaitement compatibles avec l'idée que les jugements intuitifs *sont* des activations de programmes affectifs, c'est-à-dire des réactions émotionnelles. Une classe particulière d'états mentaux – par exemple la prise de conscience qu'une souffrance est causée de manière intentionnelle plutôt que de manière non intentionnelle – déclenche une réaction émotionnelle typique qui, à son tour, cause une évaluation morale consciente (pour une analyse similaire, voir Haidt & Joseph 2004). De plus, en cas de réactions émotionnelles conflictuelles, certaines réactions peuvent être systématiquement plus prégnantes que d'autres. Par exemple, le fait de causer du tort à autrui par le biais d'un contact physique peut déclencher une réaction émotionnelle plus forte que le fait de causer le même tort de manière indirecte. Si un tort est *ressenti* comme plus important que l'autre, cela influencera les jugements moraux consciemment défendus. Une régularité dans la force de nos réactions émotionnelles face à certaines classes de situations peut donner l'impression que les jugements évaluatifs sont gouvernés par des règles intuitives. Mais cette illusion est due à une tendance à focaliser notre attention sur les évaluations *exprimées* par les gens ; c'est exactement la manière dont Mikail, Hauser et collègues procèdent. Mais si l'on pense plutôt en termes de réactions émotionnelles différenciées, on peut expliquer ces règles apparentes sans devoir postuler un sens moral spécialisé et gouverné par des règles intuitives.

Une telle analyse qui accorde un rôle important aux émotions se trouve confirmée par des expériences récentes menées sur des sujets normaux et des patients souffrant de lésions du cortex frontal et ventromédian – c'est-à-dire des patients présentant des déficiences au niveau affectif (Ciaramelli *et al.* 2007 ; Koenigs *et al.* 2007). Au

quotidien, ces patients présentent des comportements moraux anormaux et un manque notoire d'intérêt pour les règles morales ; en revanche, leurs capacités intellectuelles sont préservées. Les résultats d'expériences basées sur le dilemme du trolleybus ont montré que même si ces patients se montrent parfaitement capables de distinguer entre les situations moralement et non-moralement pertinentes,<sup>7</sup> leurs jugements au sujet des situations morales sont biaisés par rapport aux sujets normaux. Probablement à cause de leurs déficiences émotionnelles, les patients considèrent les deux scénarios du trolleybus mentionnés plus haut – pousser une manette ou précipiter une personne sur les voies – comme également acceptables. Ces études semblent montrer que les réactions émotionnelles sont des ingrédients essentiels au moins pour certains jugements.

#### **4. Le processus d'évaluation rapide**

Dans ce qui suit, je présente en détail mon tableau affectif à la lumière des résultats et analyses de la section précédente. Je m'inspire également d'écrits philosophiques sur les émotions. Commençons avec le premier volet du tableau : le processus évaluatif affectif.

Lorsque nous sommes témoin de certaines situations nous formons des « évaluations affectives spontanées ». On pourrait les considérer comme des jugements primitifs (de Sousa 1987; Döring 2007). Plus précisément, ces évaluations sont des réactions émotionnelles composées de deux parties intrinsèquement liées l'une à l'autre : une évaluation rapide non inférentielle<sup>8</sup> d'un état de fait, et un état affectif particulier lié à une sensation particulière. En utilisant le terme de « sensation », je voudrais exprimer l'idée que les réactions émotionnelles possèdent une certaine phénoménologie dans notre expérience personnelle qui est intimement liée à l'évaluation.<sup>9</sup>

---

<sup>7</sup> Cela est sans doute dû au fait qu'au cours de leur vie, avant que leur maladie ne se déclare, ces patients ont appris à faire ce genre de distinctions.

<sup>8</sup> Pour une analyse du caractère non inférentiel de l'évaluation émotionnelle, voir Sabine Döring (2007).

<sup>9</sup> En philosophie des émotions, il y a actuellement un grand débat autour de la question de savoir si l'aspect affectif de la réaction émotionnelle est causé par l'aspect évaluatif (Robinson 2005) ou si les deux aspects se confondent en un tout indissociable (Goldie 2000 ; Döring 2007). Pour mon propos, il ne me paraît

Par son aspect évaluatif, une réaction émotionnelle possède un contenu intentionnel – au sens où elle est dirigée vers un objet.<sup>10</sup> Elle est une manière de prêter une attention sélective à certaines caractéristiques d'une situation et de les percevoir d'une certaine manière (voir de Sousa 1987). Dans un contexte moral, une évaluation affective spontanée est une manière primitive et non réflexive d'interpréter une situation comme des « devant être réalisée » ou « devant être évitée ». Cette évaluation s'exprime par un état affectif qui implique des sensations négatives ou positives – comme celles liées au fait de se sentir moralement dégoûté ou de se sentir bienveillant.

Cette manière d'interpréter le monde dépend de différents facteurs : certaines tendances génétiquement innées, nos expériences passées et l'environnement social dans lequel nous nous trouvons. Voyons dans le détail comment cela fonctionne. Dans une certaine mesure, la manière dont nous évaluons les faits moralement pertinents dépend de systèmes simples et rapides dont les résultats, après réflexion, nous satisfont généralement assez bien – quoique ce ne soit pas toujours le cas (Lazarus 1991; Ekman, 1999). Grâce à ces raccourcis mentaux, nous avons plus ou moins les mêmes réactions émotionnelles face à des situations similaires. Par exemple, nous considérons généralement que les actions qui causent un tort à autrui sont moralement plus condamnables que les omissions qui causent un tort à autrui ;<sup>11</sup> ou de manière plus

---

cependant pas indispensable de prendre position dans ce débat. Il me paraît suffisant de dire qu'une réaction émotionnelle est à la fois affective et évaluative et que les deux aspects sont intimement liés l'un à l'autre.

<sup>10</sup> Dans la mesure où les réactions émotionnelles ont un contenu intentionnel, on peut dire qu'elles sont cognitives. Toutefois, cela n'implique pas que les réactions émotionnelles se réduisent à des jugements propositionnels, c'est-à-dire à des sortes de croyances (voir Döring, 2007). Cela n'implique pas non plus – comme semblent le penser certains (Döring 2007, Goldie 2000, Tappolet 2000) – que les émotions nous fournissent un accès épistémique à des valeurs existantes. En fait, cette idée est assez problématique. Comme le souligne Kevin Mulligan (2007), la connaissance n'est normalement pas conçue comme une *réaction*. C'est la raison pour laquelle il est difficile de concevoir les émotions ou les affects comme des producteurs de connaissance. Je ne poursuivrai pas cette discussion ici puisque mon projet n'est pas de déterminer ce qui est bon et mal, pas plus que d'interroger la fiabilité ou la justesse des réactions émotionnelles.

<sup>11</sup> Il vaut la peine de remarquer que ce principe fonctionne même dans les cas où l'action causerait moins de dommages que l'omission (à ce propos, voir Baron 1998 ; Sunstein 2005).

générale, nous blâmons systématiquement les actions qui causent de la souffrance chez autrui (Nichols 2004) de même que les comportements non loyaux ou opportunistes. Il semblerait donc qu'il y a un ensemble de mécanismes ou de capacités spécialement conçues pour générer des évaluations rapides au sujet de ce qui est acceptable ou non.

Au-delà de ces aspects rigides, nos facultés émotionnelles sont plastiques et ouvertes, au sens où elles se développent en fonction de nouvelles expériences. Cela signifie qu'elles sont influencées par nos expériences passées et par le contexte dans lequel nous nous trouvons.<sup>12</sup> En d'autres termes, toute réaction émotionnelle est particulière, car imbriquée dans la vie d'un sujet avec toutes ses contingences culturelles et personnelles. Considérons un exemple. Martial est grondé plusieurs fois par ses parents pour avoir caché les jouets de sa petite sœur. Disons que ces reproches l'ont mis mal à l'aise. Suite à cela, Martial associe de manière inconsciente le fait de subir des reproches avec le fait de se sentir mal à l'aise. Et puisque cette sensation est déplaisante, son cerveau marquera – au sens de la théorie des marqueurs somatiques de Damasio (2001/1994) – les reproches ainsi que les situations susceptibles de les déclencher comme étant « à éviter ». Dans ce contexte, on peut parler de mécanismes appris qui déclenchent une sensation émotionnelle qui à son tour guide les futures évaluations et comportements. Dans notre exemple, chaque fois que Martial éprouve le désir d'une action susceptible de déclencher le reproche de ses parents – par exemple, cacher les jouets de sa sœur –, son cerveau reconstituera le marqueur somatique du malaise – même si c'est de manière moins vive que dans les circonstances réelles – et cela le motivera à se restreindre d'agir de la sorte.<sup>13</sup>

Nos évaluations affectives spontanées dépendent également des divers motifs que nous avons par ailleurs. Par exemple, des recherches en psychologie sociale ont montré que les gens sont profondément influencés par le désir de maintenir des relations sociales plaisantes. Ce motif guide leurs attitudes évaluatives et la manière dont ils

---

<sup>12</sup> La proportion dans laquelle les émotions sont encodées génétiquement ou le résultat de l'éducation ou de la culture est une question très débattue quoique mal posée. Il n'est pas possible d'y apporter une réponse satisfaisante car chaque cas est différent.

<sup>13</sup> De plus, comme nous le verrons plus loin, au cours du processus réflexif, Martial attribuera probablement une valeur négative à ce type d'actions.

traitent l'information (Chen *et al.* 1996). De manière plus générale, beaucoup d'études ont également été menées sur le phénomène de la contagion émotionnelle, la tendance à ressentir des émotions similaires à celles d'autrui (Simner 1971 ; Hatfield *et al.* 1994). En résumé, notre environnement social dirige nos réactions émotionnelles dans le sens de la consistance avec celles de nos voisins.

Avant de conclure cette section, il me paraît important de préciser encore deux points. Premièrement, même si nous sommes passifs dans nos réactions émotionnelles – au sens où elles s'imposent à notre conscience de manière soudaine, incontrôlée et sans effort –, cela n'empêche pas que ces réactions émotionnelles puissent découler de pensées et croyances complexes ; elles peuvent résulter de la prise de conscience de spécificités très fines d'une situation ; elles peuvent même être produites au cours d'une réflexion rationnelle. De plus, leur grain particulier dépendra en bonne partie de nos capacités intellectuelles et du contexte culturel dans lequel nous avons grandi car les émotions complexes sont des choses que nous apprenons et qui sont largement forgées par la culture.

Deuxièmement, même si les réactions émotionnelles sont des *évaluations*, on peut se demander s'il est pertinent de les considérer comme des *jugements*. Les réactions émotionnelles nous fournissent des inputs pour le raisonnement et la formulation de jugements moraux conscients. En un sens, on peut dire que ce sont des dispositions à croire. Toutefois, si l'on est enclin à considérer les jugements comme des expressions verbales, les réactions émotionnelles ne pourraient pas être incluses dans cette catégorie.

## **5. Les jugements de valeur spontanés**

Une réaction émotionnelle est généralement exprimée ; elle ne reste pas tacite. Il y a deux manières de le faire. La première est de l'exprimer sous forme non verbale. C'est le cas par exemple lorsque notre faciès se contracte et exprime la colère, ou lorsqu'on crie « Ah ! ». Cela correspond plus ou moins à la manière dont les émotivistes conçoivent les jugements moraux (Ayer 1946/1936). Mais cette manière d'exprimer une évaluation affective spontanée reste encore dans le domaine de l'incontrôlable ; l'expression n'est que l'extension automatique et directe de la réaction émotionnelle

interne et s'inscrit dans une polarité positive – approbation – ou négative – désapprobation.

La deuxième manière d'exprimer une évaluation affective spontanée comporte une expression verbale. C'est-à-dire que l'on peut conceptualiser les réactions émotionnelles en termes de « c'est horrible ! » ou « c'est bien ! ». Il ne s'agit pas uniquement d'une expression d'approbation ou de désapprobation – comme l'expression non verbale – mais d'un énoncé conscient et réflexif par le moyen duquel on attribue une valeur à quelque chose. Il semblerait donc qu'il s'agisse d'une sorte de reconstruction cognitive de la réaction émotionnelle. Cet apport cognitif, qui s'ajoute à la forme primitive du jugement de valeur, permet en fait la production d'un nouveau jugement légèrement plus complexe : ce que je vais appeler un « jugement de valeur spontané ». L'idée est qu'en mettant des mots sur une évaluation affective spontanée, nous ne nous contentons pas de l'exprimer mais produisons en fait un nouveau jugement. De cette manière, nous faisons un premier pas en direction d'un processus réflexif dans le cadre duquel il s'agit de trouver de bonnes raisons pour nos réactions émotionnelles et pour les jugements, normes et valeurs que nous sommes enclins à accepter.

Toutefois, même s'ils sont légèrement plus raffinés – car nous ne nous contentons pas d'exprimer une attitude mais attribuons une valeur à une situation –, ces jugements spontanés demeurent relativement confus et s'effectuent également de manière largement automatique. En produisant un jugement spontané, nous affirmons sans plus de précision que quelque chose est bon ou mauvais parce que nous le ressentons de cette manière.<sup>14</sup> Sur une échelle de complexité, les jugements spontanés se situent donc entre deux extrêmes : la simple expression d'approbation ou de désapprobation et l'assertion d'un énoncé normatif à caractère objectif.

Il semblerait que ce soit ce genre de jugements que les psychologues comme Haidt ou Hauser testent dans leurs expériences lorsqu'ils focalisent l'attention sur les jugements rapides et intuitifs. Il vaut la peine de remarquer ici que les philosophes de la morale s'intéressent généralement aux jugements sophistiqués – ceux que l'on peut

---

<sup>14</sup> A ce niveau de mon explication, je ne voudrais pas défendre une position volontariste qui accorderait un rôle trop important à la volonté dans la production des jugements spontanés.

justifier – plutôt qu'aux jugements spontanés décrits dans cette section. Les psychologues de la morale en revanche, considèrent les deux sortes de jugements sans réellement les distinguer et les traitent plus ou moins de la même manière. Je doute de la pertinence d'une telle approche et favoriserais au contraire un modèle explicatif qui différencie clairement les deux sortes de jugements. De ce point de vue, je pense que mon tableau affectif aide à réconcilier les résultats des études empiriques avec les approches philosophiques, précisément parce qu'il introduit une distinction entre les jugements spontanés et sophistiqués. Ces derniers font l'objet du second volet de mon tableau affectif auquel est consacrée la prochaine section.

## 6. Le processus réflexif

Cette section rend compte du versant hautement cognitif de l'activité morale ; c'est ici qu'entrent en jeu les « jugements de valeur sophistiqués ».

Nous nous trouvons souvent confrontés à des désaccords entre nos jugements évaluatifs spontanés et ceux produits par autrui, ou entre nos jugements et le comportement d'autrui. Parce que notre survie dépend largement de notre capacité de mener une vie cohérente et coordonnée à celle de nos voisins, nous éprouvons un double besoin : d'un côté, nous voulons nous prouver à nous-mêmes et à autrui la pertinence de nos réactions émotionnelles, de l'autre côté nous désirons qu'autrui partage nos réactions émotionnelles. Ces deux objectifs nous incitent à nous engager dans une activité complexe au cours de laquelle nous réfléchissons sur les raisons qui justifient nos réactions émotionnelles et jugements spontanés – c'est ce que j'appelle le « processus réflexif » – et mettons en pratique de manière plus ou moins consciente diverses méthodes pour influencer les réactions et convictions de nos voisins.

Dans le cadre du processus réflexif, nous cherchons à justifier nos jugements spontanés. Nous pouvons le faire de manière minimale en nous demandant simplement s'ils sont appropriés à la situation ; il s'agit alors de savoir si tous les aspects pertinents ont été pris en compte, si notre engagement personnel a altéré notre jugement, etc. Nous pouvons également – et c'est ce qui est le plus intéressant en matière de morale – chercher à leur donner une justification forte au sens de *fondement* ; dans ce cas, nous faisons reposer nos jugements sur des normes et des valeurs. Dès lors, pour les besoins

de notre tableau, il serait intéressant de disposer d'une explication de la manière dont nous définissons et choisissons les normes et valeurs auxquelles nous adhérons. Je vais argumenter en faveur d'un modèle tripartite selon lequel il y a trois déterminants pour nos normes et valeurs : les affects – et en particulier les émotions –, les considérations rationnelles, et la persuasion sociale.

Il semblerait que nous choisissons nos normes et valeurs en grande partie en fonction des sensations qui sont causées en nous dans des situations concrètes. Si un état de choses cause en nous une forte sensation désagréable, dès que nous y réfléchissons, nous aurons tendance à attribuer une valeur négative à cet état de choses – et inversement pour les sensations positives (de Sousa 1987 ; Hookway 2002). Ce lien peut s'établir de deux façons : la première met en jeu les sensations qui font partie d'une réaction émotionnelle alors que la seconde met en jeu des sensations simples.

Les réactions émotionnelles, c'est-à-dire les évaluations affectives spontanées, peuvent révéler ce que nous valorisons inconsciemment. En effet, il semble qu'une partie de ce que nous valorisons nous est rendu épistémologiquement accessible précisément par la médiation des réactions émotionnelles (voir Goldie 2000, pp. 48-49). Par exemple, si nous sommes horrifiés à la vue de notre voisin qui bat son enfant pour le plaisir – et par là produisons une évaluation affective spontanée –, nous aurons tendance à attribuer une valeur négative à ce type de comportement et à établir une norme correspondante qui interdit ce genre d'action. Selon cette explication, notre esprit est déjà imprégné de manière inconsciente par un certain nombre de valeurs et de règles et les réactions émotionnelles servent à nous en faire prendre conscience. Toutefois, ce n'est certainement pas la seule manière de définir nos valeurs. Si c'était le cas, nous naîtrions avec un dispositif de valeurs prédéterminé et notre seul horizon serait de les découvrir.

Le choix d'une valeur peut aussi résulter d'une simple sensation associée à une représentation conceptuelle de la cause de cette sensation. Souvenons-nous de Martial ; après s'être fait gronder par ses parents, il éprouve un certain malaise chaque fois qu'il conçoit l'idée de cacher les jouets de sa sœur. Cette sensation de malaise associée à un certain type de situation l'incitera à valoriser négativement la situation en question ; il pensera que c'est mal de cacher les affaires de sa sœur. En bref, nous avons tendance à

attribuer une valeur négative à ce qui cause en nous des sensations désagréables – et inversement pour les valeurs positives.<sup>15</sup> De plus, les valeurs et normes que nous choisissons sous l'influence de simples sensations peuvent même supplanter des valeurs et normes induites par des réactions émotionnelles préalablement ancrées en nous. Prenons un exemple parlant : Prosper prône la fidélité à tel point qu'il est dégoûté à l'idée même d'un individu qui trompe son partenaire. Mais un jour il rencontre Célestine et craque... il trompe sa femme avec elle. Cette expérience s'avère si plaisante qu'il cesse d'éprouver de l'aversion envers les partenaires infidèles.<sup>16</sup> L'exemple de Prosper et Célestine illustre également un autre principe : la puissance des sensations joue un rôle dans la manière dont nous choisissons nos normes et valeurs ; plus une sensation est forte, plus nous avons tendance à attribuer une valeur à ce qui en est la cause.

Cette analyse est confirmée par des données empiriques. Les études mentionnées plus haut sur les patients souffrant de lésions au cortex préfrontal et ventromédian (Ciaramelli *et al.* 2007 ; Koenigs *et al.* 2007) montrent que les jugements de valeur dépendent de notre expérience affective et de notre compréhension des émotions d'autrui. L'étude a révélé que les patients présentant des déficiences au niveau

---

<sup>15</sup> Sabine Döring (2007) défend un point de vue similaire mais basé sur une compréhension des émotions comme perceptions. Selon elle, « en faisant l'expérience d'une émotion, le monde apparaît au sujet comme s'il était tel que l'émotion le lui représente ». Elle ajoute que « les émotions jouent un rôle dans le raisonnement avant que le raisonnement ne joue son rôle dans la rationalisation d'une action » – cela dit, je m'éloigne de la position de Döring lorsqu'elle prétend que les émotions permettent de percevoir des valeurs extérieures (à ce propos, voir note 10).

<sup>16</sup> Je ne voudrais pas affirmer par là que tous les individus confrontés à une situation similaire réviseraient leurs valeurs après avoir rencontré une Célestine. Certains pourraient admettre que leur comportement est moralement condamnable et qu'ils souffrent de faiblesse de la volonté. D'autres pourraient maintenir leur aversion morale face à l'infidélité mais dans le même temps se trouver des raisons spécifiques pour justifier leur comportement. Mon intention avec cet exemple est plutôt d'illustrer le fait que nos émotions exercent une influence importante sur nos choix normatifs, même si un effort de réflexion rationnelle permet d'atténuer leurs effets.

émotionnel considèrent comme moralement acceptable de précipiter un homme sur les voies du trolleybus...<sup>17</sup>

Voilà pour une première explication de la manière, probablement la plus courante, de choisir nos normes et valeurs. Il existe cependant d'autres manières de les définir. Nous pouvons nous lancer dans des raisonnements d'inférence plus complexes et, partant de normes et valeurs préalablement acceptées, en déduire de nouvelles. C'est à ce stade que la pensée rationnelle prend sa place dans l'activité morale. Toutefois, sans parler du fait que tout processus d'inférence part de prémisses données, il faudrait se garder d'y accorder trop d'importance (Haidt 2003). Comme le font bien remarquer Joshua Greene et Jonathan Haidt (2002), il est vrai que les gens s'engagent souvent dans des débats moraux réels ou fictionnels mais la plupart du temps, ces efforts sont dirigés vers le renforcement ou la transmission de jugements, valeurs ou normes sur lesquels les sujets ont déjà fixé leur choix par avance.

Le troisième déterminant des normes et valeurs est la persuasion sociale. Dans les contextes sociaux, les gens tentent de s'influencer mutuellement et d'assurer un consensus avec leurs amis ou alliés. Il existe même des mécanismes psychologiques – auxquels est consacrée toute une littérature empirique – qui régissent et assurent l'efficacité de cette influence mutuelle. Selon les anthropologues évolutionnistes Joseph Henrich, Robert Boyd et collègues, les êtres humains sont largement influencés dans leurs jugements et pratiques par un certain nombre de biais psychologiques. Les biais sont des sortes de règles d'apprentissage social du type 'Copie qui a le plus de succès !', ou 'Copie la majorité !' Un des biais les plus influents est celui du conformisme : les êtres humains ont tendance à reproduire les comportements les plus fréquents de la population dans laquelle ils évoluent (Henrich & Boyd 1998). Un autre biais est celui du prestige : les êtres humains tendent à prendre pour modèle des individus qui paraissent avoir du succès ou qui semblent posséder des qualités ou des connaissances supérieures (Henrich & Gil-White 2001). En ayant recours aux techniques de la théorie des jeux ou de la simulation par ordinateur, ces chercheurs tentent de déterminer les conditions

---

<sup>17</sup> Pour des données similaires, voir Adina Roskies (2003 ; 2006). Cette chercheuse a pu montrer que les patients ventromédians – ceux qui souffrent de déficits au niveau des réactions émotionnelles – peuvent produire des jugements moraux même en l'absence de réactions émotionnelles.

d'adaptation évolutionnaire de ces mécanismes psychologiques et émettent l'hypothèse qu'ils ont évolué parce qu'ils permettent aux individus de bénéficier à peu de frais des avantages liés à l'adoption d'un comportement, d'une norme, d'une valeur ou d'une coutume. Chandra Sripada et Stephen Stich (2005, pp. 150-155 ; voir aussi Fessler & Navarrete 2003) renforcent à l'aide d'exemples concrets cette hypothèse de l'existence de biais psychologiques.<sup>18</sup> Faisant référence à différentes études empiriques, ils montrent que les êtres humains sont souvent incapables de juger correctement les avantages et désavantages de variantes culturelles – par exemple dans le domaine des innovations ou des tabous liés à la nourriture – et prennent leurs décisions sous l'influence des biais du conformisme et du prestige.<sup>19</sup> De plus il a été montré dans différentes expériences psychologiques menées par Jody Davis et Caryl Rusbult (2001) que nous sommes directement influencés dans nos jugements par ceux de nos amis, alliés ou proches parents. Cela confirme selon eux l'existence du phénomène d'« alignement d'attitude » (*attitude alignment*), une tendance, chez les partenaires d'interaction, à modifier leurs attitudes respectives de manière à ce qu'elles convergent (pour plus de données, voir Haidt 2001).<sup>20</sup>

En bref, nous choisissons systématiquement des valeurs et normes congruentes avec nos émotions, lesquelles sont partiellement innées et partiellement acquises au moyen de l'apprentissage individuel au grès des interactions avec notre environnement social. Cela n'empêche en rien l'émergence de conflits à l'intérieur même de nos

---

<sup>18</sup> Allan Gibbard (1990), un des rares philosophes de la morale qui prennent les données empiriques réellement au sérieux, est un précurseur d'un bon nombre d'idées développées dans cet article. Entre autres choses, il intègre dans son explication de l'activité morale, des mécanismes psychologiques comme l'« influence normative » (*normative influence*) – similaire au biais du prestige – ou l'exigence de cohérence (*demand for consistency*) – similaire à l'émotion épistémique qui sera introduite plus loin dans cette section.

<sup>19</sup> Ces auteurs soulignent également le fait que malgré leur fonction de révélateurs d'adaptations, ces biais peuvent mener à la propagation d'innovations et pratiques hautement maladaptives si elles sont prises isolément ; c'est par exemple souvent le cas en matière de tabous alimentaires.

<sup>20</sup> Précisons également que les mécanismes psychologiques mentionnés sont étroitement liés aux émotions. Par exemple, le biais du prestige se manifeste par une certaine sensibilité émotionnelle face à des personnes que l'on considère comme prestigieuses ; il induit la tendance à copier les pratiques, et adopter leurs croyances et valeurs.

productions morales – c'est-à-dire nos actions, les valeurs ou normes auxquelles nous souscrivons, les jugements de valeur spontanés ou sophistiqués que nous produisons – ou entre nos productions morales et celles de nos voisins. C'est ici que l'activité rationnelle refait surface dans le tableau affectif : c'est elle qui attire notre attention sur ce genre d'incohérences. Cette prise de conscience déclenche en nous une émotion épistémique<sup>21</sup> que l'on peut appeler « demande de cohérence » et se caractérise par un sentiment d'inconfort.<sup>22</sup> Ce sentiment nous inclinera à tenter de rétablir la cohérence (pour des études empiriques supportant cette idée, voir Haidt 2001 ; Moskowitz *et al.* 1999 ; Thagard 1992). Nous ne pourrons pas nous empêcher de procéder à des raisonnements inférentiels, à ouvrir notre cœur à nos sentiments les plus profonds et à chercher l'avis de notre entourage jusqu'à ce que l'harmonie soit rétablie. Enfin, l'activité rationnelle signifie également la prise en considération des questions d'ordre pratique et l'orientation du choix de nos normes en fonction de leur réalisabilité et des autres buts que nous nous sommes fixés.

Les grandes lignes de mon tableau affectif sont maintenant tracées. Ce tableau mériterait évidemment d'être développé davantage mais la présentation qui en a été faite est suffisante pour permettre de prendre position dans le débat entre l'internalisme et l'externalisme. Dans ce qui suit, je commencerai par défendre l'idée que la motivation morale est toujours déclenchée par les réponses affectives avant de prendre position dans le vieux débat philosophique.

## 7. L'hypothèse humienne affective

L'approche humienne de la motivation est une conception philosophique qui a déjà une longue histoire. Selon cette théorie, la motivation morale dépend toujours d'un état conatif préexistant, lequel est habituellement conçu en termes de désir.<sup>23</sup> Je pense

---

<sup>21</sup> Certains préféreront parler de mécanisme psychologique (Gibbard 1990).

<sup>22</sup> En effet, promouvoir à la fois p et -p, nous met dans une situation analogue à un dilemme pratique où il est impossible de réaliser tous les buts que nous nous sommes fixés.

<sup>23</sup> Notons que l'approche humienne n'impose aucune prise de position particulière au niveau du débat entre l'internalisme et l'externalisme. Même si beaucoup de philosophes d'obédience humienne

qu'il y a de bonnes raisons d'accepter une telle analyse de la motivation à la différence qu'au lieu des désirs, je suggérerai de mettre la focale sur les *états affectifs* – dans leur forme simple ou en tant qu'ils font partie intégrante de phénomènes plus complexes comme les épisodes émotionnels par exemple ; au fond, les désirs motivent précisément parce qu'ils sont intrinsèquement liés à des états affectifs. Appelons cette position révisée une « hypothèse humienne affective ». Cette dernière semble d'ailleurs plus proche de la position défendue par Hume lui-même lorsqu'il écrit : « Il paraît évident que les fins ultimes des actions humaines ne peuvent jamais, en aucun cas, être expliquées par la raison, mais qu'elles se recommandent entièrement aux sentiments et aux affections des hommes, sans dépendre aucunement des facultés intellectuelles. » Hume 1751, Appendix I, V). Selon Hume, les considérations rationnelles ne peuvent en elles-mêmes générer ni des affects ni des actions. Dans cette section, j'argumenterai en faveur de cette idée.

L'hypothèse humienne affective semble la meilleure explication possible pour une série de données empiriques issues de la psychologie. Jennifer Beer et collègues (2003 ; 2006) ont mené des études comparatives entre des personnes en bonne santé et des patients souffrant de dommages cérébraux au niveau du cortex orbitofrontal, la partie du cerveau connue pour être liée aux processus émotionnels. En faisant jouer un certain nombre de paramètres, les expérimentateurs ont pu mettre en évidence une corrélation nette entre les troubles de régulation du comportement social et l'absence d'émotions liées à la conscience de soi (*self-conscious emotions*) comme la honte, l'embarras ou la fierté. Il semblerait donc que la capacité d'avoir des émotions liées à la conscience de soi dans des circonstances appropriées est un facteur essentiel à la régulation du comportement social. Dans la même veine, différentes études sur des psychopathes ou patients avec des lésions des parties ventromédiane et préfrontale du cerveau (Damasio 2001/1994, Blair *et al.* 2005, Bechara *et al.* 1996, Gray *et al.* 2003) révèlent que ces patients présentent un dysfonctionnement de leur système affectif alors même qu'ils ne montrent aucun déficit dans leurs raisonnements pratiques. D'un côté, ils sont parfaitement capables de penser en termes de normes, d'un autre côté en revanche, ils

---

défendent une position externaliste, cela ne doit pas nécessairement être le cas. Michael Smith (2006/1994) par exemple est un humien internaliste.

montrent peu d'intérêt pour les normes sociales et morales et se comportent de manière antisociale – à un degré nettement plus dramatique dans le cas des psychopathes.

Ces données empiriques révèlent l'existence d'une corrélation forte qui relève probablement d'un lien essentiel entre les processus affectifs et l'action morale. De ce point de vue ces données semblent plus compatibles avec une analyse humienne de la motivation qu'avec une approche rationaliste selon laquelle les croyances morales suffisent à nous motiver à agir (Nagel 1970; Dancy 2000). Plus précisément, ces données soutiennent l'hypothèse humienne *affective* parce que l'explication la plus simple et directe pour les importantes corrélations observées est de considérer les processus affectifs comme des ingrédients nécessaires à la prise de décision. Ce n'est bien sûr pas un argument décisif mais voici une raison supplémentaire de favoriser cette approche. Le philosophe Matteo Mameli (2005) a analysé de manière détaillée les résultats des expériences de Damasio sur des patients présentant des lésions préfrontales. Il constate qu'en plus d'être incapables d'agir de manière sociale, ces patients ont également beaucoup de peine à prendre des décisions pratiques de tous les jours, malgré le fait qu'ils soient tout à fait capables de mener des raisonnements logiques.<sup>24</sup> Si, comme le suggèrent ces données, les considérations rationnelles ne sont pas suffisantes pour motiver à prendre des décisions, on peut se demander quel est l'ingrédient manquant. Mameli pense que c'est précisément parce que les patients ont des déficiences au niveau du système affectif qu'ils éprouvent des difficultés à prendre des décisions. Il en déduit que nos capacités générales à prendre des décisions pratiques, y compris les décisions morales, dépendent de manière cruciale du bon fonctionnement de notre système affectif ; en d'autres termes, nos décisions pratiques sont le produit de nos émotions. Ainsi il écrit : « Chez les êtres humains, le choix entre différentes actions (...) est toujours déterminé par les réactions émotionnelles [*emotional feelings*] causées par la prise en considération des résultats probables des différentes actions possibles, et non par une froide analyse des coûts et bénéfices » (Mameli 2005, p. 171, ma

---

<sup>24</sup> Il semblerait qu'en plus de leurs déficits affectifs, les psychopathes présentent des difficultés de langage – ils peinent à catégoriser les notions sémantiques abstraites – d'attention ou d'orientation (Kiehl 2006). Toutefois, les psychopathes sont habituellement capables de raisonnements inférentiels. C'est tout ce dont nous avons besoin pour donner du crédit à l'analyse de Mameli.

traduction). Un corrélat important de cette explication est que même si les énoncés moraux résultent de processus d'inférence conscients, le choix des actions n'est pas dû à ce processus. En effet, un processus d'inférence est indépendant de l'affect ; ce n'est rien de plus qu'une succession de croyances qui suivent certaines règles. Dans les termes de Mameli : «les croyances au sujet de ce qui est socialement approprié dans certaines circonstances n'exercent aucune force motivationnelle sur nos prises de décision, sauf si elles peuvent déclencher des réactions émotionnelles [*emotional feelings*] qui nous motivent à agir en accord avec le contenu de nos croyances » (2005, p. 171, ma traduction). Précisons que Mameli n'affirme pas l'impossibilité d'être motivé à produire des actions dont on sait qu'elles auront pour conséquence de déclencher des sensations désagréables. Ces cas sont possibles à condition que l'on prédise pour les mêmes actions, des résultats plaisants d'un autre point de vue *et* que ces aspects plaisants l'emportent sur les aspects négatifs (Mameli 2005, p. 172).<sup>25</sup>

Considérons maintenant deux modèles concurrents à l'hypothèse humienne affective et demandons-nous s'ils sont capables de la surpasser. On pourrait défendre l'idée que les réactions morales affectives ne sont rien de plus que des épiphénomènes d'une cognition morale intuitive ou tout simplement de nos croyances morales. Selon cette approche, nous comprenons – ou du moins pensons comprendre – ce qui est la bonne chose à faire ; cette connaissance ou simple croyance nous motive à agir *et* déclenche habituellement une réaction émotionnelle. Le problème d'une telle interprétation réside dans son coût théorique : elle soulève plus de questions qu'elle n'en résout. En effet, emprunter cette voie argumentative implique de devoir trouver un nouvel éclairage pour d'autres aspects du phénomène à expliquer. Si la cognition ou croyance morale est suffisante pour motiver, pourquoi nous sentons-nous souvent outrés ou coupables ? Si ces émotions se réduisent à des épiphénomènes, pourquoi sont-elles si largement partagées dans l'espèce humaine ? Quel rôle jouent-elles si ce n'est pas un rôle motivant ? Pourquoi certains types d'émotions – par exemple, la peur des serpents – sont clairement responsables de nos actions alors que d'autres types d'émotions –

---

<sup>25</sup> Dans la même veine, Joshua Greene (2005) passe en revue une série de recherches qui montrent que l'efficacité des normes sociales dépend d'émotions suffisamment fortes pour motiver à l'action.

celles qui sont supposées épiphénoménales – sont dénuées d'une telle efficacité causale ?

Une autre alternative – sans doute plus intéressante – serait d'affirmer que les croyances morales sont motivantes parce qu'elles sont capables, en elles-mêmes – sans autres médiateurs –, de générer des désirs (Dancy 2000) ou des réactions émotionnelles qui, à leur tour, génèrent des actions. Mais cette stratégie explicative soulève également de nouvelles questions. Selon cette approche, il devrait exister un processus par le biais duquel les croyances morales causent l'activation de désirs ou de mécanismes affectifs qui, ultimement, nous incitent à agir. Mais cette connexion causale n'apparaît pas dans tous les cas de figure ; les croyances au sujet de la beauté d'un objet ou de la validité d'un argument par exemple ne génèrent aucun désir particulier. On peut donc légitimement se demander pourquoi précisément les croyances morales génèrent des désirs ou des réactions émotionnelles.

De manière plus problématique encore, les deux modèles alternatifs décrits ci-dessus se trouvent confrontés à la difficulté d'expliquer pourquoi les patients souffrant de lésions du cortex ventromédian sont apparemment capables de penser qu'ils devraient agir d'une certaine manière sans pour autant être motivés à le faire (Roskies 2003 ; Damasio *et al.* 1990).<sup>26</sup>

En fin de compte, le principe de parcimonie s'impose de lui-même et incite à adopter l'hypothèse humienne affective : une réponse affective est une cause nécessaire à l'action, c'est le seul ingrédient capable de disposer les êtres humains à agir.

## **8. Retour au débat sur l'internalisme *versus* externalisme**

L'hypothèse humienne affective, combinée avec mon tableau affectif, permet de tirer une série de conclusions intéressantes dans le contexte du débat sur l'internalisme

---

<sup>26</sup> On pourrait objecter que les patients souffrant de lésions du cortex ventromédian ne sont pas capables de reporter correctement leurs propres états mentaux ; soit ils se trompent au sujet de leurs véritables jugements ou alors ils n'endossent pas sincèrement leurs assertions morales (Cholbi 2006) ; mais ces interprétations ont été réfutées de manière convaincante par Adina Roskies (2006).

et l'externalisme. Comme nous allons le voir, la balance penchera finalement du côté d'une position hybride entre l'internalisme et l'externalisme.

L'application de l'hypothèse humienne au tableau affectif met en valeur deux caractéristiques importantes des jugements de valeurs. Premièrement les jugements spontanés sont intrinsèquement motivants puisque ce sont des expressions de réactions émotionnelles. Plus précisément, en raison du fait que les jugements spontanés sont intrinsèquement liés à des états affectifs – les réactions émotionnelles –, ils sont un facteur clef pour l'action morale. Deuxièmement les produits du processus réflexif – les valeurs et normes acceptées ainsi que les jugements sophistiqués – ne sont pas motivants en eux-mêmes puisqu'ils ne possèdent aucune dimension émotionnelle. La motivation doit provenir d'une source externe ; soit d'une réaction émotionnelle corrélée au jugement sophistiqué, soit d'un désir d'obéir au jugement en question – par exemple parce que nous voulons éviter la punition ou soigner nos relations.

Cette analyse fournit une explication simple des cas de conflit entre nos jugements moraux et nos actions. Elle permet de considérer la possibilité de juger qu'une chose devrait être faite sans pour autant être motivé à le faire. Une telle situation ne peut cependant surgir qu'en cas d'élaboration froide d'un jugement sophistiqué qui ne correspond pas à l'évaluation émotionnelle de la situation. Souvenons-nous de l'exemple de Raymond qui refuse son aide à une personne en difficulté. Il se peut que Raymond ne ressente rien de particulier dans cette situation, même s'il est conscient du fait que son attitude est moralement condamnable au vu des principes moraux auxquels il adhère. Une telle situation peut par exemple être due au fait que Raymond est extrêmement fatigué ce jour là ou au fait que son esprit est occupé par la triste perspective de son divorce agendé au lendemain...

L'analyse proposée ici permet de remettre en question la thèse philosophique internaliste. Si mon analyse est correcte, les jugements moraux et les considérations morales en général ne motivent pas forcément à l'action ; à l'exemple de Raymond, il est possible pour un agent de produire sincèrement un jugement moral sophistiqué sans pour autant être incliné à le suivre. D'un autre côté, l'externalisme est également affaibli au niveau des jugements de valeur spontanés puisque l'hypothèse humienne combinée au tableau affectif nous présentent ces jugements comme *nécessairement* motivants.

Au premier abord, cette position hybride peut s'avérer assez dérangeante puisqu'elle met en évidence l'inefficacité causale des jugements moraux paradigmatiques, ceux qui peuvent être fondés sur des normes et des valeurs. Dès lors, on pourrait penser que l'action moralement vertueuse est toujours produite par hasard, si bien que l'on ne peut par vraiment parler d'action *morale*. Mais il s'agit là d'une conclusion trop pessimiste. D'une part, les actions vertueuses sont intrinsèquement liées à des jugements évaluatifs spontanés. D'autre part, les liens entre nos réactions automatiques et notre activité réflexive sont extrêmement forts. De fait, la plupart de nos normes, valeurs et jugements sophistiqués sont intimement corrélés à nos réactions émotionnelles parce que ces dernières influencent largement nos évaluations réflexives.<sup>27</sup> En outre, l'effet inverse n'est pas négligeable : les produits du processus réflexif peuvent exercer un impact indirect sur nos réactions émotionnelles. Par exemple, en affinant notre compréhension cognitive de la situation jugée, nous pouvons déceler de nouvelles caractéristiques pertinentes qui vont modifier nos réactions émotionnelles ou en déclencher de nouvelles. De manière encore plus intéressante, je pense qu'il faut prendre au sérieux la possibilité de manipuler les mécanismes émotionnels ancrés dans nos esprits, ceux qui causent nos réactions émotionnelles – à l'exemple de l'expérience du dégoût hypnotique mentionnée plus haut. Par exemple, si nous voulons qu'une de nos normes ou valeurs ait une force motivante – ou du moins soit corrélée à nos motivations concrètes –, nous pouvons tenter de forger un mécanisme émotionnel correspondant. Dans une certaine mesure nous manipulons déjà nos émotions sans même avoir à nous fixer consciemment ce but. Malheureusement, ces questions sont extrêmement complexes et délicates, si bien que je ne pourrai pas les développer plus avant dans le cadre de cet article. Quoiqu'il en soit, en prenant conscience de l'enchevêtrement complexe des relations entre les processus affectifs et réflexifs, on comprend qu'il n'y a aucune raison de déprécier les actions qui découlent des premiers.

---

<sup>27</sup> C'est d'ailleurs la raison pour laquelle l'internalisme motivationnel est convaincant au premier abord : les jugements sophistiqués semblent motivants parce qu'ils sont habituellement corrélés avec des réactions émotionnelles congruentes – mais en réalité, seules ces dernières sont les vecteurs de nos actions morales.

## 9. Conclusion

Dans cet article, j'ai détaillé les deux volets d'un tableau affectif représentant la manière dont les gens jugent et se comportent moralement. Ce tableau est largement inspiré d'études empiriques. Je n'ai pas cherché à adopter une attitude philosophique de confrontation vis-à-vis de ces études. Plutôt que de chercher des raisons de douter de la vérité ou de la pertinence de leurs conclusions, je suis partie du présupposé que ces analyses empiriques sont en mesure de nous dire quelque chose d'intéressant au sujet de la manière dont nous pensons et agissons moralement et peuvent être intégrées dans mon tableau affectif. Une tâche future serait bien sûre de renforcer ce tableau à l'aide de données empiriques supplémentaires et d'arguments de nature philosophique. En attendant, j'espère avoir pu relever de manière convaincante deux effets positifs des processus affectifs dans le contexte de l'activité morale. Premièrement, ils ne dénaturent pas nos jugements mais les façonnent de manière directe – dans le cas des jugements spontanés – et indirecte – dans le cas des jugements sophistiqués. Deuxièmement, ils sont responsables de notre comportement moral ; les jugements de valeur spontanés sont intrinsèquement motivants parce qu'ils résultent directement de nos réactions émotionnelles. De plus, le tableau de l'activité morale proposé dans cet article permet d'expliquer pourquoi les normes auxquelles nous souscrivons et les jugements sophistiqués qui leurs sont associés ne prescrivent pas toujours ce que nous choisissons en fin de compte de faire ; ils échouent à nous motiver lorsqu'ils ne sont pas soutenus par les jugements de valeur spontanés.\*

\* Mes remerciements vont à Otto Bruun, Julien Deonna, Ronald de Sousa, Julien Dutant, Curzio Chiesa, Chloe FitzGerald, Jean Gayon, Peter Goldie, Daniel Schulthess, et Fabrice Teroni pour avoir pris le temps de critiquer et discuter les idées présentées dans cet article.

## References

Alfred J. Ayer: *Language, truth, and logic* (London: V. Gollancz, 1946).

- Jonathan Baron: *Judgment misguided: Intuition and error in public decision making* (New York: Oxford University Press, 1998).
- Antoine Bechara, Hanna Damasio, Daniel Tranel, Antonio R. Damasio: *Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex*, in *Cerebral Cortex* 6 (1996) pp. 215-25.
- Jennifer S Beer, Oliver P John, Donatella Scabini, Robert T. Knight: *Orbitofrontal cortex and social behavior: Integrating self-monitoring and emotion-cognition interactions*, in *Journal of Cognitive Neuroscience* 18 (2006) pp. 871-79.
- Jennifer S. Beer, Erin A. Heerey, Dacher Keltner, Donatella Scabini, Robert T. Knight: *The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage*, in *Journal of Personality and Social Psychology* 85 (2003) pp. 594-604.
- James Blair, Derek Robert Mitchell, Karina Blair: *The psychopath: Emotion and the brain* (Malden: Blackwell, 2005).
- David Owen Brink: *Moral realism and the foundations of ethics* (Cambridge; New York: Cambridge University Press, 1989).
- Serena Chen, David Shechter, Shelly Chaiken: *Getting at the truth or getting along: accuracy- versus impression-motivated heuristic and systematic processing*, in *Journal of Personality and Social Psychology* 71 (1996) pp. 262-75.
- Michael Cholbi: *Belief attribution and the falsification of motive internalism*, in *Philosophical Psychology* 19 (2006) pp. 607-16.
- Elisa Ciaramelli, Michela Muccioli, Elisabetta Ladavas, Giuseppe di Pellegrino: *Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex*, in *Social Cognitive and Affective Neuroscience* 2 (2007) pp. 84-92.
- Fieri Cushman, Liane Young, Marc Hauser: *The role of conscious reasoning and intuition in moral judgment: testing three principles of harm*, in *Psychological Science* 17 (2006) pp. 1082-89.
- Antonio R. Damasio: *L'erreur de Descartes: la raison des émotions* (Paris: Odile Jacob, 2001 / 1995).
- Antonio R. Damasio, Daniel Tranel, Hanna Damasio: *Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli*, in *Behavioural Brain Research* 41 (1990) pp. 81-94.
- Jonathan Dancy: *Practical reality* (Oxford; New York: Oxford University Press, 2000).
- Jody L Davis, Caryl E. Rusbult: *Attitude alignment in close relationships*, in *Journal of Personality and Social Psychology* 81 (2001) pp. 65-84.
- Ronald de Sousa: *The rationality of emotion* (Cambridge: MIT Press, 1990).
- Sabine A. Döring: *Seeing what to do: Affective perception and rational motivation*, in *Dialectica* 61 (2007) pp. 363-94.
- Paul Ekman: *Basic Emotions*, in *Handbook of cognition and emotion*, ed. T. Dalgleish, M.J. Power (Chichester; New York: Wiley, 1999), pp. 45-60.
- Daniel M. Fessler, Carlos David Navarrete: *Meat is good to taboo: Dietary proscriptions as a product of the interaction of psychological mechanisms and social processes*, in *Journal of Cognition and Culture* 3 (2003) pp. 1-40.
- Allan Gibbard: *Wise choices, apt feelings: A theory of normative judgment* (Cambridge: Harvard University Press, 1990).

- Peter Goldie: *The emotions: A philosophical exploration* (Oxford; New York: Oxford University Press, 2000).
- Nicola S. Gray, Malcolm J. MacCulloch, Jennifer Smith, Mark Morris, Robert J. Snowden: *Forensic psychology: Violence viewed by psychopathic murderers*, in *Nature* 423 (2003) pp. 497-8.
- Joshua D. Greene, *Cognitive Neuroscience and the Structure of the Moral Mind*, in *The Innate Mind: Structure and Contents*. ed. P. Carruthers, et al. Oxford, New York: Oxford University Press (2005), pp. 338-52.
- Joshua D. Greene, Jonathan Haidt: *How (and where) does moral judgment work?*, in *Trends in Cognitive Sciences* 6 (2002) pp. 517-23.
- Joshua D. Greene, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, Jonathan D. Cohen: *An fMRI investigation of emotional engagement in moral judgment*, in *Science* 293 (2001) pp. 2105-08.
- Jonathan Haidt: *The emotional dog and its rational tail: A social intuitionist approach to moral judgment*, in *Psychological Review* 108 (2001) pp. 814-34.
- Jonathan Haidt, Craig Joseph: *Special Issue on Human Nature*, in *Deadalus* (2004) pp. 55-66.
- Elaine Hatfield, John T. Cacioppo, Richard L. Rapson: *Emotional contagion* (Cambridge, New York, Paris: Cambridge University Press, 1994).
- Marc D. Hauser: *Moral minds: How nature designed our universal sense of right and wrong* (New York: Ecco, 2006).
- Joseph Henrich, Robert Boyd: *The evolution of conformist transmission and the emergence of between-group differences*, in *Evolution and Human Behavior* 19 (1998) pp. 215-41.
- Joseph Henrich, Robert Boyd: *Why people punish defectors. Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas*, in *Journal of Theoretical Biology* 208 (2001) pp. 79-89.
- Joseph Henrich, Francisco J. Gil-White: *The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission*, in *Evolution and Human Behavior* 22 (2001) pp. 165-96.
- Christopher Hookway: *Emotions and epistemic evaluations*, in *The cognitive basis of science*, ed. P. Carruthers, et al. (Cambridge; New York: Cambridge University Press, 2002), pp. 251-62.
- David Hume: *Enquête sur les principes de la morale* (Paris: Flammarion, 1991 / 1751).
- William James: *What is an emotion?*, in *Mind* 9 (1884) pp. 188-205.
- Kent A. Kiehl: *A cognitive neuroscience perspective on psychopathy: Evidence for paralimbic system dysfunction*, in *Psychiatry Research* 142 (2006) pp. 107-28.
- Richard S. Lazarus: *Emotion and adaptation* (New York: Oxford University Press, 1991).
- Michael Koenigs, Liane Young, Ralph Adolphs, Daniel Tranel, Fiery Cushman, Marc Hauser, Antonio Damasio. *Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements*, in *Nature* 446 (2007), pp. 908-1011.
- Matteo Mameli: *The role of emotions in ecological and practical rationality*, in *Emotion, evolution, and rationality*, ed. D. Evans, P. Cruse (Oxford; New York: Oxford University Press, 2005 / 2004), pp. 159-78.
- John Mikhail: *Aspects of the theory of moral cognition: Investigating intuitive knowledge of the prohibition of intentional battery and the principle of double*

- effect*, in *Social Science Research Network* (2005) pp. 1-129, <http://papers.ssrn.com/sol3/DisplayAbstractSearch.cfm> (15.12.2007).
- Gordon B. Moskowitz, Ian Skurnik, Adam D. Galinsky: *The history of dual process notions, and the future of pre-conscious control*, in *Dual-process theories in social psychology*, ed. S. Chaiken, Y. Trope (New York: Guilford Press, 1999), pp. 12-36.
- Kevin Mulligan: *Intentionality, knowledge and formal objects*, in *Hommage à Wlodek* (2007) pp. 1-17, <http://www.fil.lu.se/HommageaWlodek/> (15.12.2007).
- Thomas Nagel: *The possibility of altruism* (Oxford: Clarendon P., 1970).
- Shaun Nichols: *Sentimental rules: On the natural foundations of moral judgment* (Oxford: Oxford University Press, 2004).
- Robert J. Nisbett, Timothy D Wilson: *Telling more than we can know: Verbal reports on mental processes*, in *Psychological Review* 84 (1977) pp. 231-59.
- Jenefer Robinson: *Deeper than reason: Emotion and its role in literature, music, and art* (Oxford: Oxford University Press, 2005).
- Adina Roskies: *Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy"*, in *Philosophical Psychology* 16 (2003) pp. 51-66.
- Adina Roskies: *Patients with ventromedial frontal damage have moral beliefs*, in *Philosophical Psychology* 19 (2006) pp. 617-27.
- Marvin L. Simner: *Newborn's response to the cry of another infant*, in *Developmental Psychology* 5 (1971) pp. 136-50.
- Michael Smith: *The moral problem* (Oxford; Cambridge: Blackwell, 1995).
- Chandra Sripada, Stephen P. Stich: *Evolution, culture and the irrationality of the emotions*, in *Emotion, evolution, and rationality*, ed. D. Evans, P. Cruse (Oxford; New York: Oxford University Press, 2005 / 2004), pp. 133-58.
- Cass R. Sunstein: *Moral heuristics*, in *Behavioral and Brain Sciences* 28 (2005) pp. 531-42.
- Sigrún Svavarsdóttir: *Moral cognitivism and motivation*, in *The Philosophical Review* 108 (1999) pp. 161-219.
- Christine Tappolet: *Emotions et valeurs* (Paris: Presses universitaires de France, 2000).
- Paul Thagard: *Conceptual revolutions* (Princeton: Princeton University Press, 1992).
- Thalia Wheatley, Jonathan Haidt: *Hypnotically induced disgust makes moral judgments more severe*, in *Psychological Science* 16 (2005) pp. 780-84.