

Internalisme et externalisme : Le problème de la motivation morale

François Jaquet & Florian Cova

Ceci est une version pré-épreuves. Merci de citer la version publiée :
Jaquet, F. & Cova, F. (2018). Internalisme et externalisme : Le problème de la motivation morale. In O. Desmons, S. Lemaire & P. Turmel (eds.), *Manuel de métaéthique*. Hermann.

1. Introduction : motivation morale et métaéthique

Imprégné par une éducation religieuse, Gédéon considère que la prière du soir est moralement obligatoire. Dieu le dit dans la bible, et Il ne ment jamais. Les convictions de Gédéon sont bien connues de ses proches, auxquels il les expose régulièrement. Loin de s'en agacer, ceux-ci trouvent ses interventions plutôt divertissantes.

Simone juge la consommation de viande injuste. Cette pratique implique une grande souffrance pour les animaux, que ne saurait justifier le maigre plaisir qu'elle nous procure en échange. Parce qu'elle a suivi un cours intitulé « L'éthique pour les nuls », Simone sait par ailleurs que l'argument de la tradition ne vaut rien.

Compte tenu des jugements moraux auxquels ils souscrivent respectivement, tout porte à croire que Gédéon effectue une prière chaque soir et que Simone ne mange pas de viande. Si l'on découvrait que Gédéon ne prie que le matin ou que Simone mange du poulet, on les taxerait probablement d'hypocrisie. « Bien que Gédéon *affirme* que la prière du soir est obligatoire, il ne *pense* pas que la prière du soir est obligatoire », dirait-on, et : « Bien que Simone *affirme* que la consommation de viande est injuste, elle ne *pense* pas que la consommation de viande est injuste ». Plus généralement, nous nous attendons en principe à ce que nos congénères soient disposés à agir sur la base de leurs jugements moraux.

L'*internalisme motivationnel* (ci-après, « internalisme ») légitime cette attente, puisqu'il soutient qu'il existe une connexion nécessaire entre les jugements moraux et la motivation. Selon la variante la plus simple de cette théorie, un sujet qui juge un acte moralement bon ou obligatoire sera nécessairement motivé à l'accomplir, tandis qu'un sujet qui juge un acte moralement mauvais ou interdit sera nécessairement motivé à s'en abstenir (pour peu qu'il en soit l'agent). L'internalisme s'oppose à l'*externalisme*, qui nie l'existence d'une telle connexion systématique entre les jugements moraux et la motivation. Ainsi, bien que la variante la plus simple de l'externalisme concède que nous sommes généralement motivés par nos jugements moraux, elle insiste sur la possibilité, pour un sujet, de juger qu'un acte est moralement obligatoire sans pour autant être motivé à l'accomplir et de juger qu'un acte est moralement mauvais sans pour autant être motivé à s'en abstenir. Alors que la connexion entre jugement moral et motivation est nécessaire selon l'internalisme, elle est contingente selon l'externalisme.

Si l'internalisme séduit certains philosophes, c'est principalement en vertu de deux implications qu'il a en métaéthique. La première concerne le débat millénaire

qui porte sur les raisons que nous avons d'agir moralement. Depuis l'Antiquité¹, les philosophes se demandent s'il est rationnel de faire ce qui est moralement bon et de ne pas faire ce qui est moralement mauvais. Les problèmes que soulève cette question se cristallisent tout particulièrement dans le personnage de l'amoraliste. Imaginez un individu qui, bien qu'il saisisse parfaitement la distinction entre le bien et le mal et partage même vos convictions morales, ne semble pas le moins du monde disposé à se conduire moralement. Lorsque vous mettez en cause sa conduite en lui signalant que ce qu'il fait est mal, il se contente de hausser les épaules : ce qu'il fait est immoral, et alors ? L'amoraliste est-il irrationnel ? Y a-t-il une raison pour lui de se conduire moralement, mais dont il ne tient pas compte ? Et, plus généralement, à quoi bon se conduire moralement ?

Les philosophes antiques, de Platon aux Stoïciens, surmontent cette difficulté en supposant deux choses : (i) tout être raisonnable désire avoir une vie bonne, et (ii) la vertu morale est une composante nécessaire de la vie bonne. Ainsi, il serait dans l'intérêt bien compris de l'amoraliste de se conduire moralement, car c'est là la seule façon d'obtenir ce qu'il est rationnel pour lui de désirer. Malheureusement, cette solution repose sur une thèse substantielle au sujet de la vie bonne que peu de philosophes acceptent de nos jours. Dans la lignée de Kant, la philosophie contemporaine tend en effet à dissocier bien-être et vertu morale.

L'internalisme a ceci d'intéressant qu'il permet d'esquiver ce débat. En effet, à supposer que les jugements moraux soient nécessairement motivants, si un sujet n'est pas motivé à faire ce qu'il dit qu'il devrait faire, c'est qu'il ne croit pas ce qu'il dit. En d'autres termes, il s'ensuit de l'internalisme qu'il n'y a rien à répondre à l'amoraliste tel que défini ci-dessus, parce qu'il est impossible de n'être pas motivé par ses jugements moraux. Du point de vue internaliste, la question « À quoi bon accomplir cette bonne action ? » n'a simplement aucun sens.

L'autre conséquence intéressante de l'internalisme concerne un débat central en métaéthique, qui porte sur la nature des jugements moraux. Sur ce terrain, deux camps s'affrontent. Les cognitivistes soutiennent que les jugements moraux sont des états mentaux cognitifs (plus précisément, des croyances) : juger que la prière du soir est moralement obligatoire, c'est *croire* qu'elle est obligatoire, se la *représenter* comme obligatoire. A contrario, les non-cognitivistes prétendent que les jugements moraux sont des états mentaux conatifs (autrement dit, des attitudes affectives, favorables ou défavorables) : juger que la consommation de viande est injuste, c'est simplement y être opposé, ce qui n'implique pas de se la représenter comme injuste. Puisque les attitudes conatives ne prétendent pas représenter le monde, elles ne sont à proprement parler ni vraies ni fausses. Le non-cognitivismisme a donc une implication singulière : les jugements moraux sont eux-mêmes dénués de valeur de vérité.

Combiné à une conception très populaire de la motivation, l'internalisme fait pencher la balance en faveur du non-cognitivismisme. Selon la théorie dite « huméenne » de la motivation, les états cognitifs ne sont motivants qu'une fois associés à des états conatifs. Votre croyance que le dernier épisode de la saga Star Wars est sorti ne vous motivera à vous rendre au cinéma que si vous désirez le voir ; en l'absence d'un tel désir, elle ne s'opposera pas à ce que vous fassiez plutôt une sieste. Plus généralement, d'après la théorie huméenne, en l'absence

¹ Platon, *La République*, Paris, Garnier Flammarion, 2002.

d'attitudes conatives correspondantes, les croyances demeurent inertes du point de vue de la motivation.

Voici l'un des principaux arguments en faveur du non-cognitivism. Pour tout sujet S et pour tout acte A,

- (1) Nécessairement, si S juge que A est moralement bon, S est motivé à accomplir A.
- (2) Nécessairement, si S est motivé à accomplir A, S est dans un certain état conatif.
- (3) Donc, nécessairement si S juge que A est moralement bon, S est dans un certain état conatif².

La première prémisse de cet argument n'est autre que la version simple de l'internalisme que nous avons introduite ci-dessus. Quant à la seconde, elle découle directement de la théorie huméenne de la motivation. Ensemble, elles entraînent logiquement la conclusion.

À supposer, comme l'affirme cette dernière, qu'il est impossible de juger qu'une action est bonne (ou mauvaise) sans être dans un état conatif, il y a fort à parier que les jugements moraux *sont* les états conatifs en question, plutôt que des croyances. Si l'internalisme simple est vrai, il est par conséquent probable que le non-cognitivism le soit également. Compte tenu de l'importance du débat qui oppose les cognitivistes aux non-cognitivistes, il vaut donc la peine de s'arrêter sur les raisons qui plaident en sa faveur et sur celles qui s'y opposent.

Comme nous allons toutefois le constater, il n'est pas clair du tout que l'internalisme soit vrai. Dans un premier temps, nous distinguerons un certain nombre de ses variantes et écarterons celles d'entre elles qui sont moins directement intéressantes pour les grands débats métaéthiques. Nous examinerons ensuite trois arguments strictement philosophiques pour ou contre l'internalisme, qui échouent invariablement à trancher la question. Enfin, nous présenterons des arguments empiriques – relevant respectivement de la psychologie, des neurosciences et de la philosophie expérimentale – dont nous verrons qu'ils ont de meilleures chances d'y parvenir.

2. Variantes de l'internalisme simple et autres formes d'internalisme

Tel que nous l'avons défini, l'internalisme est la thèse selon laquelle, pour tout sujet S et pour tout acte A,

- (1) Nécessairement, si S juge que A est moralement bon, S est motivé à accomplir A.

Cet énoncé est assez simple. Comme nous allons maintenant le constater, il est néanmoins susceptible de plusieurs interprétations, lesquelles génèrent autant de variantes de l'internalisme.

Un premier éclaircissement concerne le statut épistémique de la thèse internaliste. L'internalisme est-il supposé exprimer une vérité a priori, conceptuelle, ou une vérité a posteriori, métaphysique ? L'internalisme conceptuel suit la première interprétation : il est vrai en vertu de la nature même du concept de jugement moral (c'est-à-dire, par définition) que former un jugement moral suppose d'être motivé à s'y conformer. Par conséquent, celui ou celle qui attribue

2 À l'inverse, si un sujet juge un acte moralement mauvais, il sera motivé à s'en abstenir et, par voie de conséquence, aura une attitude conative.

un jugement moral à un sujet qui n'est pas motivé à s'y conformer est conceptuellement confus. L'internalisme métaphysique opte quant à lui pour la seconde lecture : les jugements moraux sont nécessairement accompagnés d'états motivationnels seulement en ce sens qu'ils le sont toujours et dans tous les mondes possibles. Celui ou celle qui attribue un jugement moral à un sujet qui n'est pas motivé à s'y conformer n'est pas forcément conceptuellement confus ; il se peut qu'il ignore simplement une loi métaphysique.

Les philosophes se sont largement concentrés sur la variante conceptuelle de l'internalisme³. L'internalisme métaphysique est moins clairement pertinent du point de vue métaéthique, parce qu'il est compatible avec le cognitivisme. Il se pourrait en effet que les jugements moraux soient des croyances qui s'accompagnent toujours et dans tous les mondes possibles d'états mentaux conatifs, ce qui rendrait vrais tant l'internalisme métaphysique que le cognitivisme. À l'inverse, s'il est conceptuellement inconcevable qu'un sujet forme un jugement moral sans être motivé à s'y conformer, il semble que les jugements moraux soient eux-mêmes motivants. Il est alors nettement plus tentant de nier qu'ils soient cognitifs (sauf, évidemment, à rejeter la théorie huméenne de la motivation).

Une deuxième distinction concerne la notion de motivation impliquée dans la thèse internaliste. En quel sens serions-nous *motivés* par nos jugements moraux ? On distingue communément la motivation *pro toto* de la motivation *pro tanto*. Marc est *pro toto* motivé à aller voir *Le réveil de la force* s'il est disposé à le faire *toute chose considérée* : à coup sûr, il le fera si l'occasion se présente. En revanche, Marc est *pro tanto* motivé à aller voir ce film s'il est disposé à le faire *dans une certaine mesure* – il peut être *pro tanto* motivé à aller au cinéma, même s'il désire plus intensément encore faire la sieste. Cette distinction donne lieu à deux formes d'internalisme. D'après l'internalisme fort, nous sommes *pro toto* motivés à agir conformément à nos jugements moraux, tandis que nous sommes seulement *pro tanto* motivés à le faire selon l'internalisme faible. À de rares exceptions près⁴, les philosophes ont davantage défendu la version faible de l'internalisme⁵.

Les thèses internalistes que nous avons rencontrées jusqu'ici sont inconditionnelles : elles n'envisagent pas que la connexion entre les jugements moraux et la motivation puisse être défaite. Il existe cependant des variantes conditionnelles de l'internalisme, qui ont la forme suivante (pour tout sujet S, tout acte A et une condition C) :

³ Richard Mervyn Hare, *The language of morals*, Oxford, Oxford University Press, 1952.

⁴ Selon le prescriptivisme (voir *Ibid.*), les énoncés moraux entraînent des impératifs, lesquels expriment des désirs. Les désirs en question sont alors *pro toto*. En effet, si l'impératif « Brosse-toi les dents ! » exprime un désir, ce dernier ne sera pas *pro tanto*, puisqu'il est contradictoire de dire : « Brosse-toi les dents et ne te brosse pas les dents ! » bien que l'on puisse sans incohérence désirer *pro tanto* à la fois que son enfant se brosse les dents et qu'il ne se brosse pas les dents. L'internalisme fort s'ensuit donc du prescriptivisme.

⁵ Simon Blackburn, *Ruling passions: a theory of practical reasoning*, New York, Oxford University Press, 1998 ; Michael Smith, *The Moral Problem*. Oxford: Wiley-Blackwell, 1994.

- (4) Nécessairement, si S juge que A est moralement bon, alors S est motivé à accomplir A, pourvu que C soit remplie.

L'internalisme conditionnel est plus modeste que l'internalisme simple, parce qu'il admet des cas de jugements moraux non motivants, des cas où la condition C n'est pas satisfaite. Bien que nécessaire, la relation entre le jugement moral et la motivation est alors conçue comme conditionnelle.

L'internalisme conditionnel suppose alors d'identifier la condition C qui est pertinente – telle que nécessairement, lorsqu'elle est satisfaite, les jugements moraux sont motivants. Jusqu'ici, trois propositions ont été avancées, générant trois variantes de l'internalisme conditionnel : C est satisfaite (i) lorsque l'auteur du jugement est psychologiquement normal⁶, (ii) lorsqu'il est pratiquement rationnel⁷ ou (iii) lorsque son sens moral fonctionne correctement⁸. La première condition est violée si le sujet est atteint d'un trouble mental, tel que l'apathie, la dépression ou la psychopathie ; la seconde, s'il souffre de faiblesse de la volonté ; la troisième s'il est victime d'un trouble d'ordre perceptuel : contrairement aux sujets normaux, il ne perçoit pas (bien) les propriétés et les faits moraux.

Par souci d'exhaustivité, mentionnons une dernière théorie. Selon l'internalisme collectif, les jugements moraux sont nécessairement connectés à la motivation non pas *de leur auteur* mais des *membres de sa communauté*. Néanmoins, si un sujet forme un jugement moral, il s'ensuit nécessairement que la plupart des membres de sa communauté sont motivés à agir sur la base de leurs jugements moraux⁹.

Bien qu'elles aient généré une multitude de publications, les versions collectives et conditionnelles de l'internalisme ne sauraient être mobilisées à l'appui du non-cognitivism. Si une condition doit être remplie pour garantir la motivation morale, et que les sujets déprimés ou irrationnels soient en mesure de former des jugements moraux sans être motivés, on peut douter que ces derniers soient des états conatifs.

Quelles sont alors les raisons d'accepter ou de rejeter l'internalisme ?

3. Les arguments conceptuels

Que ce soit en faveur de l'internalisme ou à son encontre, les arguments ne manquent pas. Certains sont proprement philosophiques, qui relèvent essentiellement de l'analyse conceptuelle et ne reposent que marginalement sur des données empiriques. Nous leur dédions la présente section. La suivante sera consacrée aux approches plus expérimentales.

a. L'internalisme explique la motivation morale

Un premier argument en faveur de l'internalisme s'appuie sur le constat que nous sommes généralement motivés à agir conformément à nos jugements moraux. Compte tenu de leurs convictions morales respectives, nous nous attendons à ce que Gédéon prie tous les soirs et à ce que Simone soit végétarienne. L'internalisme

⁶ Blackburn, *Ruling passions: a theory of practical reasoning*.

⁷ Smith, *The Moral Problem*.

⁸ David McNaughton, *Moral vision: an introduction to ethics*, Oxford, Blackwell, 1988.

⁹ Matthew Bedke, « Moral judgments purposivism: saving internalism from amorality », *Philosophical Studies*, 144, 2009, p. 189-209.

nous fournit une explication extrêmement simple, et par conséquent attrayante, de cette observation : souscrire à un jugement moral, c'est *déjà* être motivé. Puisque Gédéon juge qu'il faut prier tous les soirs et que Simone juge que manger de la viande est injuste, il n'est pas surprenant qu'il effectue une prière tous les soirs et qu'elle soit végétarienne.

Mais les externalistes ne sont pas à court d'explications. Nous en citerons trois, dont voici la première. Certains psychologues affirment que nos jugements moraux résultent plus souvent de nos émotions que de nos raisonnements¹⁰. Témoins d'un acte cruel, nous ne jugeons pas dans un premier temps que l'acte est cruel pour ensuite en inférer qu'il est injuste ; l'acte nous révolte, en conséquence de quoi nous le jugeons injuste. Or, il est bien connu que les émotions ont un autre effet : celui de nous motiver à agir. Nous serons naturellement disposés à ne pas commettre d'actes qui nous révoltent. Motivation et jugement moraux pourraient donc bien constituer les deux effets d'une même cause – un état affectif –, ce qui expliquerait leur corrélation. Pour autant, cette corrélation ne serait pas conceptuellement nécessaire.

Il se pourrait en outre que nous soyons motivés à agir moralement simplement parce que nous désirons faire ce qui est moralement bon. Cette explication appelle toutefois la prudence, car l'expression « désirer faire ce qui est moralement bon » est notoirement ambiguë. Elle peut être interprétée soit de manière *de re*, soit de manière *de dicto*. Selon la première lecture, lorsque nous désirons faire ce qui est moralement bon, nous désirons accomplir une action donnée, qui se trouve par ailleurs être bonne. Selon la seconde interprétation, quand nous désirons faire ce qui est moralement bon, nous désirons accomplir l'action bonne, quelle qu'elle soit, précisément parce qu'elle est bonne. Imaginons que vous fassiez don d'une partie de votre argent à une association qui vient en aide aux enfants malades du cancer. Si vous êtes motivé à le faire par votre seul désir d'améliorer leur sort, indépendamment du fait que cela soit juste, bon ou moralement admirable, alors votre désir est *de re* : il porte sur la *chose* même que vous faites (aider les enfants) et non sur ce qu'on pourrait en *dire* moralement (faire ce qui est bon ou juste). À l'inverse, si vous aidez ces enfants non parce que vous êtes sensibles à leur sort mais précisément et seulement parce que c'est là votre devoir, alors votre désir est *de dicto* : il ne porte pas sur la *chose* même que vous faites (aider les enfants) mais sur la façon dont cet acte peut être *qualifié* moralement (faire ce qui est moralement bon ou juste). Bien entendu, il est possible d'être motivé par les deux types de désirs en même temps (vous pouvez être motivé à la fois parce que vous souhaitez les aider *et* parce que voulez faire ce qui est moralement bon), mais il n'en reste pas moins important de les distinguer.

La deuxième explication externaliste recourt au désir *de dicto* de faire le bien¹¹. Ainsi, il se pourrait que Gédéon fasse ce qu'il juge moralement obligatoire (prier tous les soirs) simplement parce qu'il désire faire ce qui est moralement obligatoire, *quoi que cela implique*, et qu'il juge par ailleurs que prier tous les soirs est obligatoire. Dans ce cas, la motivation morale de Gédéon ne s'expliquerait pas

¹⁰ Jonathan Haidt, « The emotional dog and its rational tail: a social intuitionist approach to moral judgment », *Psychological review*, 108(4), 2001, p. 814-834.

¹¹ Sigrun Svavarsdottir, « Moral cognitivism and motivation », *The Philosophical Review*, 108(2), 1999, p. 161-219.

(seulement) par sa croyance que prier tous les soirs est obligatoire, mais par son désir de faire ce qui est moralement obligatoire dans la simple mesure où cela est moralement obligatoire (c'est-à-dire son désir *de dicto* de faire ce qui est obligatoire). En l'absence de ce désir supplémentaire d'accomplir l'action qui est obligatoire en tant qu'elle est obligatoire, Gédéon aurait jugé que prier tous les soirs est obligatoire sans pour autant être motivé à le faire. Parce que la présence de ce désir est un fait contingent, cette explication fait de la motivation morale de Gédéon un fait contingent. Gédéon est de fait motivé à agir en accord avec son jugement moral, mais il aurait tout aussi bien pu ne pas l'être, conformément à l'externalisme.

La troisième explication externaliste fait appel aux désirs *de re* de faire ce qui est moralement obligatoire¹². Il se pourrait que Simone fasse ce qu'elle juge moralement obligatoire (s'abstenir de manger de la viande) simplement parce qu'elle n'aime pas le goût de la viande, indépendamment de son jugement moral. En dépit de leur corrélation, le jugement moral de Simone n'expliquerait pas sa motivation : il se trouverait simplement que Simone juge que ce qu'elle désire faire en l'occurrence est moralement obligatoire.

S'il faut reconnaître qu'aucune de ces explications ne s'applique à tous les cas, on admettra néanmoins qu'elles se complètent harmonieusement. Ensemble, elles ne paraissent pas moins convaincantes que l'explication internaliste. Le premier argument ne permet donc pas d'établir que les jugements moraux sont nécessairement motivants.

Les internalistes disposent cependant d'un deuxième argument, qui s'en inspire.

b. L'internalisme explique la motivation morale en cas de conversion

Nous nous attendons non seulement à ce que nos congénères agissent en conformité avec leurs opinions morales, mais aussi à ce qu'ils modifient leur comportement (ou du moins à ce qu'ils soient, dans une certaine mesure, disposés à le faire) lorsqu'ils changent d'opinion morale. Sous l'emprise de l'enthousiasme communicatif de Simone, Gédéon s'est inscrit à « L'éthique pour les nuls ». À l'issue de la première séance, consacrée à l'éthique animale, il a radicalement changé d'avis : jusqu'alors convaincu de la légitimité de sa consommation de viande, il la juge désormais moralement condamnable. Gédéon adopte par conséquent un régime végétarien.

Les internalistes disposent d'une explication simple et élégante pour ce phénomène : souscrire à un jugement moral, c'est être motivé à agir ; rien d'étonnant donc à ce que notre motivation s'aligne sur nos jugements moraux quand nous en changeons. Tant que Gédéon considérait que manger de la viande était moralement admissible, son jugement ne le disposait évidemment pas à s'en abstenir. Il a cependant acquis cette disposition avec sa nouvelle conviction. En clair, avant comme après sa conversion morale, Gédéon était nécessairement motivé à agir conformément à son jugement moral.

¹² David Owen Brink, *Moral realism and the foundations of ethics*, Cambridge, Cambridge University Press, 1989 ; Peter Railton, « Moral realism », *Philosophical Review*, 95, 1986, p. 163-207.

Selon Michael Smith¹³, les externalistes n'ont pas tant de mal à rendre compte de notre motivation morale que du fait qu'elle perdure quand nous changeons d'opinions morales. Comme nous venons de le constater, il leur est en principe loisible d'expliquer notre motivation morale de trois manières : par nos émotions morales, par notre désir *de dicto* de faire ce qui est bon et par nos désirs *de re* de faire ce qui est bon. Dans les cas de conversion morale s'ensuivant d'un raisonnement, la première explication n'est pas satisfaisante. Gédéon n'a probablement pas été particulièrement ému par son professeur d'éthique. D'après Smith, les externalistes ne peuvent pas non plus faire appel aux désirs *de re*. En effet, on voit mal pourquoi Gédéon acquerrait un désir *de re* de cesser de manger de la viande une fois convaincu que c'est ce qu'il doit faire. Si ses désirs ont changé suite à sa conversion morale, c'est donc qu'il a un désir *de dicto* de faire ce qu'il juge bon. Avant comme après sa conversion, Gédéon a toujours désiré faire ce qui est moralement bon, quoi que cela implique. Mais depuis sa conversion, il juge que manger de la viande est injuste. Combinés, son désir et son jugement le dotent d'une nouvelle motivation : il est maintenant disposé à s'abstenir de manger de la viande.

Mais Smith ne se satisfait pas de cette explication. Plus précisément, bien qu'il concède que la présence d'un désir *de dicto* puisse expliquer les changements de motivation chez certains sujets, il nie qu'elle permette d'expliquer de tels changements chez les sujets vertueux. Et pour cause : le désir *de dicto* d'accomplir l'action qui est moralement bonne *simplement* parce qu'elle est moralement bonne, sans se soucier de ce qui fait qu'elle est bonne, aurait quelque chose de fétichiste, de l'ordre d'une étrange obsession pour la morale, dont on ne saurait *a fortiori* suspecter un agent vertueux.

Afin d'établir que les désirs *de dicto* de faire ce qui est moralement bon sont fétichistes, Smith recourt à une expérience de pensée qu'il emprunte à Bernard Williams¹⁴. Un homme aperçoit deux femmes qui sont sur le point de se noyer : son épouse et une parfaite inconnue. Après un instant de délibération, il conclut qu'il devrait sauver sa femme plutôt que l'inconnue, ce qu'il s'empresse de faire. Williams reproche toutefois à son protagoniste d'entretenir « une pensée de trop » : constatant que sa femme est en danger, il devrait immédiatement se précipiter pour la sauver, au lieu de quoi il prend le temps de délibérer. De l'avis de Smith, l'époux commettrait une erreur similaire s'il était mû par un désir *de dicto* de faire le bien. S'il est véritablement vertueux, il désirera sauver sa femme (et jugera par ailleurs que c'est là ce qu'il doit faire) ; il ne désirera pas sauver sa femme *simplement* parce qu'il juge que c'est là ce qu'il doit faire. Il aura un désir *de re* de faire ce qui est moralement bon.

Les externalistes ont offert deux réponses à cette objection. D'une part, ils ont proposé d'autres explications des changements de motivation, au-delà du désir *de dicto* de faire ce qui est moralement bon. Il est par exemple plausible qu'un agent vertueux soit par définition disposé à désirer ce qu'il juge moralement bon, et les désirs en question seront *de re*¹⁵. Pour peu qu'il soit vertueux, Gédéon sera disposé

¹³ Smith, *The Moral Problem*, p. 71-76.

¹⁴ Bernard Williams, *Moral luck*, Cambridge, Cambridge University Press, 1981.

¹⁵ David Copp, « Belief, reason, and motivation: Michael Smith's "The moral problem" », *Ethics*, 108(1), 1997, p. 33-54.

à désirer ce qu'il juge bon. Suite à sa conversion morale, il aura alors un désir *de re* de cesser de manger de la viande. Sa motivation à adopter un régime végétarien ne s'ensuivra donc pas d'une quelconque délibération, pas plus qu'il ne la dérivera d'un désir *de dicto* de bien se conduire.

D'autre part, les externalistes ont nié que le désir *de dicto* de faire le bien soit nécessairement fétichiste. L'exemple de Williams diffère de celui de Gédéon en ce qu'il ne fait pas mention d'une conversion morale : tandis que Gédéon est motivé à devenir végétarien parce qu'il a changé d'opinion quant à la moralité de la consommation de viande, les croyances morales de l'époux demeurent inchangées. Il est donc possible pour les externalistes de critiquer la motivation de l'époux sous prétexte qu'elle dérive d'un désir *de dicto* de faire le bien tout en maintenant que de tels désirs sont appropriés dans les cas de conversion morale¹⁶. Et en effet, il ne semble pas malvenu de la part de Gédéon d'être dans un premier temps motivé à arrêter de manger de la viande par un désir *de dicto* de faire le bien combiné à sa nouvelle opinion morale. Ce qui fait la vertu des agents vertueux, c'est qu'ils développent des désirs *de re* de faire le bien suite à leurs conversions morales, plutôt que de continuer d'agir sur la base de leur seul désir *de dicto* de faire le bien.

Mais certains philosophes ont même contesté que les désirs *de dicto* soient fétichistes dans les cas qui n'impliquent pas de conversion morale. Ce qui semble problématique dans le comportement de l'époux est qu'il ait besoin de délibérer pour décider de sauver sa femme. Mais, en recourant à cet exemple, Smith semble confondre les actions qui résultent d'un désir *de dicto* et celle qui résultent d'une délibération¹⁷. Certains désirs *de dicto* donnent pourtant lieu à des actions non délibérées. Ainsi, Jane fuit le tigre qui la poursuit parce qu'elle veut faire le nécessaire pour sauver sa peau. Même si elle ne délibère à aucun moment avant de fuir, son désir est *de dicto* – elle désire faire ce qu'il faut pour avoir la vie sauve, quoi que cela implique. Le bon mari pourrait aussi bien sauver sa femme parce qu'il désire faire son devoir et croit en l'occurrence devoir sauver sa femme sans délibérer une seule seconde. Le cas échéant, sa réaction n'aurait, semble-t-il, rien de problématique. Il semble donc que Smith se trompe quand il affirme que les désirs *de dicto* de faire le bien supposent d'entretenir une pensée de trop.

De toute évidence, les explications externalistes de la corrélation entre les changements d'opinion morale et de motivation chez les agents vertueux sont au moins aussi bonnes que celle que propose Smith.

c. L'amoralisme à la rescousse de l'externalisme

Parmi les objections qui ont été adressées à l'internalisme, la plus fameuse concerne un personnage que nous avons déjà rencontré. L'amoraliste, bien qu'il souscrive à des jugements moraux, ne s'en soucie pas le moins du monde lorsqu'il s'agit de passer à l'action. D'après les externalistes, nous avons l'intuition qu'il pourrait exister de tels sujets, et cette intuition est incompatible avec l'internalisme : s'il est possible de souscrire à un jugement moral sans être motivé,

¹⁶ Hallvard Lillehammer, « Smith on moral fetishism », *Analysis*, 57(3), 1997, p. 187-195.

¹⁷ Richard Joyce, *The myth of morality*, Cambridge, Cambridge University Press, 2001, p. 64-67.

il est faux que les jugements moraux soient nécessairement connectés à la motivation de leurs sujets.

La structure du débat sur l'amoralisme est maintenant bien rodée. Les externalistes lancent les hostilités en soumettant à notre sagacité un scénario dans lequel un sujet n'est absolument pas motivé par ses jugements moraux. Si nous sommes capables d'imaginer un tel sujet, c'est que l'amoralisme est conceptuellement possible et, par conséquent, que la connexion entre jugements moraux et motivation n'est pas conceptuellement nécessaire¹⁸. Mais les internalistes récusent cette description de l'amoralisme : certes, les amoralistes affirment que des actes sont mauvais sans être motivés à s'en abstenir, mais, à proprement parler, ils ne *jugent* pas que ces actes sont mauvais. En fait, les amoralistes utilisent les termes moraux comme entre guillemets : en disant qu'un acte est mauvais, ils signifient simplement que la communauté à laquelle ils appartiennent le condamne, ou devrait le faire en toute logique¹⁹. Puisque la thèse internaliste porte sur les jugements moraux plutôt que sur les locutions morales, elle s'accommode bien de la possibilité de l'amoralisme. Les externalistes objectent à cette réponse qu'elle ne prend pas leur défi suffisamment au sérieux : les amoralistes ne se contentent pas d'employer le langage moral ; ils ont aussi d'authentiques pensées morales²⁰. Les internalistes nient alors qu'il y ait là un défi à prendre au sérieux. En affirmant que les amoralistes forment de véritables jugements moraux, les externalistes commettent une pétition de principe : si l'internalisme est vrai, il est clair que les amoralistes ne sont pas capables de telles extravagances.

Ce résumé de la dialectique illustre bien l'impasse dans laquelle s'est engagée la littérature philosophique consacrée à la motivation morale. Il semble en définitive que les internalistes et les externalistes ne parlent pas de la même chose, chaque camp usant d'un concept de jugement moral qui lui est propre. Celui des internalistes implique la motivation, contrairement à celui des externalistes. Voilà tout.

4. Les arguments empiriques

Les arguments listés dans la section précédente ne sont donc pas concluants : les externalistes sont en mesure de répondre aux objections de leurs opposants, et vice-versa. Prenant acte de cet insuccès, certains philosophes sont arrivés à la conclusion qu'aborder la question de l'internalisme sous l'angle a priori était une erreur. Ils se sont alors naturellement tournés vers des approches plus empiriques, espérant trancher le débat de manière a posteriori.

À première vue, l'idée de débattre a posteriori de la vérité de l'internalisme peut paraître saugrenue, tant elle s'apparente à une erreur de catégorie. Après tout, les philosophes s'intéressent à la variante conceptuelle de l'internalisme, et si cette dernière est vraie, elle l'est a priori. La tester empiriquement n'aurait donc aucun sens. Néanmoins, tout ce qui est vrai a priori l'est également a posteriori. Bien qu'elles soient a priori, les vérités mathématiques permettent de faire des

¹⁸ Brink, *Moral realism and the foundations of ethics*.

¹⁹ Hare, *The language of morals*.

²⁰ David Owen Brink, « Externalist moral realism », *Southern Journal of Philosophy*, supplement, 1986, p. 23-42.

prédictions qu'il est possible de tester empiriquement : sachant que deux et deux font quatre, si vous mettez deux pommes et deux poires dans votre panier, il contiendra quatre fruits. Identifier un contre-exemple empirique à l'internalisme conceptuel suffirait donc à le réfuter²¹. En terres philosophiques, les contre-exemples sont des contre-exemples, qu'ils soient imaginaires ou réels.

a. L'appel à la psychologie et aux neurosciences

Parmi les données expérimentales qui sont pertinentes pour le débat qui nous occupe, certaines relèvent des sciences de l'esprit que sont la psychologie et les neurosciences. Ces données ont été mobilisées tantôt par les internalistes, tantôt par les externalistes, comme l'illustrent les trois arguments qui suivent.

a.i. Psychopathie, sociopathie acquise et amoralisme

Une manière empirique de réfuter l'internalisme consiste à identifier des cas réels d'amoralistes : des sujets qui forment d'authentiques jugements moraux, sans être pour autant motivés à s'y plier. À cet égard, les psychopathes ont rapidement attiré l'attention des philosophes empiriquement informés. Ces sujets ne semblent en effet pas se soucier du bien-être d'autrui, ni d'ailleurs des règles morales en général, à moins que les observer ne serve leurs intérêts. Ils ne conçoivent les autres que comme des moyens, qu'ils emploient au service de leurs propres fins. À première vue, les psychopathes paraissent pourtant capables de manier des concepts moraux et de comprendre les énoncés moraux. Certains chercheurs en ont conclu que ces individus ont, comme nous, des croyances morales, mais qu'agir sur la base de ces croyances ne les intéresse pas²². Les psychopathes seraient donc des contre-exemples incarnés à la thèse internaliste.

Le cas des psychopathes est cependant problématique, dans la mesure où il n'est pas certain que ces sujets soient à même de former d'authentiques jugements moraux. Premièrement, compte tenu de leur réputation de manipulateurs, il est vraisemblable que les psychopathes ne pensent pas vraiment ce qu'ils disent. C'est bien connu, ils ont tendance à nous dire ce que nous voulons entendre, et rien n'indique qu'ils gratifient les expérimentateurs d'un traitement de faveur en la matière. Ainsi, quand, lors d'une expérience, un psychopathe affirme qu'il est immoral de nuire à autrui, il se pourrait qu'il ne le fasse que dans le but de mettre l'expérimentateur en confiance. Deuxièmement, on peut envisager que les psychopathes ne comprennent pas vraiment ce qu'est une règle morale. En effet, alors que les enfants distinguent très tôt (dès l'âge de 5 ans) les règles morales (valables universellement et indépendantes d'une quelconque autorité) des règles conventionnelles (valables dans un cadre donné et dépendantes d'une autorité compétente), les psychopathes semblent avoir des difficultés à opérer cette distinction²³. Lorsqu'on leur demande pourquoi il est immoral de nuire à autrui, nombre d'entre eux en appellent à l'autorité (« c'est immoral parce que c'est

²¹ Adina Roskies, « Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy" », *Philosophical Psychology*, 16(1), 2003, p. 51-66.

²² Maaïke Cima, Franca Tonnaer et Marc D. Hauser, « Psychopaths know right from wrong but don't care », *Social Cognitive and Affective Neuroscience*, 5(1), 2010, p. 59-67.

²³ Robert James Blair, « A cognitive developmental approach to morality: investigating the psychopath », *Cognition*, 57, 1995, p. 1-29.

interdit par la loi ») plutôt qu'à l'atteinte aux droits et intérêts d'autrui (« c'est immoral parce que cela fait souffrir untel ou unetelle »). De nombreux chercheurs ont ainsi conclu que les psychopathes ne maîtrisent pas les concepts moraux fondamentaux²⁴. Pour ces raisons, on aurait tort de compter sur eux pour trancher le débat en faveur de l'externalisme.

Étant donné les problèmes que soulève le cas de la psychopathie, certains philosophes se sont penchés sur un autre trouble psychiatrique : la sociopathie acquise. L'expression « sociopathie acquise » fait référence aux troubles mentaux observés chez les individus souffrant de dommages au cortex préfrontal ventromédian suite à un accident. À l'instar de Phineas Gage, exemple historique bien connu, ces individus se distinguent par d'étranges troubles du comportement et de la prise de décision. Bien que leurs compétences intellectuelles (raisonnement, langage) demeurent intactes, ils se conduisent souvent de façon impulsive, au mépris de leur meilleur jugement. La plupart d'entre eux se lancent dans des entreprises absurdes et risquées et finissent ruinés²⁵ (Damasio, 2006). Cette déconnexion entre leurs jugements et leur comportement se retrouve dans le domaine moral : s'ils sont apparemment parfaitement capables de former des jugements moraux, les sujets atteints de sociopathie acquise se conforment rarement à ces jugements. Leurs comportements sont souvent socialement inappropriés, ce qui leur vaut la plupart du temps de perdre tous leurs amis²⁶.

La sociopathie acquise semble donc constituer un contre-exemple à l'internalisme : même s'ils ont conservé la capacité de former des jugements moraux, les sujets qui en sont atteints ont perdu toute motivation à agir en conformité avec ces jugements. Ceci étant dit, les internalistes ne peuvent-ils pas, comme dans le cas des psychopathes, rejeter cet argument en soutenant que les sociopathes ne font pas d'authentiques jugements moraux ? Certains ont adopté cette stratégie²⁷, qui est néanmoins nettement plus incertaine dans le présent contexte. En effet, jusqu'à leur accident, les patients atteints de sociopathie acquise étaient tout à fait normaux. Et l'hypothèse selon laquelle leur accident aurait affecté leur maîtrise des concepts moraux semble *ad hoc*²⁸. Il apparaît donc que les sociopathes sont en mesure de former des jugements moraux même s'ils ne sont pas motivés à s'y conformer, ce qui plaide contre l'internalisme.

Une autre stratégie consiste à soutenir que les sociopathes font un usage entre guillemets des termes moraux : lorsqu'ils affirment qu'une action est injuste, il faut comprendre qu'elle est condamnée par les normes en vigueur dans leur société. Partir du principe que les sociopathes font d'authentiques jugements moraux tout en sachant qu'ils ne sont pas motivés par ces jugements revient à

²⁴ Shaun Nichols, « How psychopaths threaten moral rationalism », *The Monist*, 85(2), 2002, p. 285-303 ; Neil Levy, « The responsibility of the psychopath revisited », *Philosophy, Psychiatry, & Psychology*, 14(2), 2007, p. 129-138.

²⁵ Antonio Damasio, *L'erreur de Descartes*, Paris, Odile Jacob, 2006.

²⁶ Roskies, « Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy" ».

²⁷ Michael Cholbi, « Belief attribution and the falsification of motive internalism », *Philosophical Psychology*, 19(5), 2006, p. 607-616.

²⁸ Adina Roskies, « Patients with ventromedial frontal damage have moral beliefs », *Philosophical Psychology*, 19(5), 2006, p. 617-627.

commettre une pétition de principe en faveur de l'externalisme²⁹. Les contre-exemples empiriques ne seraient donc pas plus satisfaisants que ceux dont nous avons parlé dans la section précédente.

a.ii. Neurosciences et externalisme

Une autre façon de rejeter empiriquement l'internalisme consiste à s'intéresser à la manière dont la motivation est implémentée dans le cerveau non pas de sujets exotiques tels que les psychopathes mais d'individus dit « normaux ». Parce que tout acte volontaire suppose le contrôle de nos mouvements par les cortex moteur et pré-moteur, et parce que ces cortex sont à leur tour sous le contrôle des ganglions de la base, Timothy Schroeder, Adina Roskies et Shaun Nichols³⁰ font de ces derniers le siège cérébral de la motivation morale. Se pose alors la question de savoir quelles aires du cerveau sont en liaison directe avec les ganglions de la base. Schroeder et ses collègues en comptent deux grands types : les aires impliquées dans la cognition (qui réalisent, entre autres choses, les croyances) et le système de récompense (qui réalise, entre autres choses, les désirs).

À première vue, cela semble indiquer que les croyances peuvent être intrinsèquement motivantes, même en l'absence de désirs. Mais ici encore, la prudence est de mise. Il n'est pas étonnant qu'il existe une voie directe entre la cognition supérieure et la motivation : après tout, la planification et la programmation des mouvements corporels fait nécessairement appel à certaines croyances au sujet de l'environnement. Toute la question est de savoir si cette voie est suffisamment directe pour qu'on puisse en conclure que les croyances sont intrinsèquement motivantes.

À cette question, Schroeder et ses collègues répondent par la négative, en s'appuyant sur la maladie de Parkinson. Chez les patients atteints de cette maladie, la liaison entre le système de récompense et les ganglions de la base se détériore, laissant seule la liaison entre cognition et motivation. Or, on observe aussi que ces patients deviennent progressivement incapables d'agir tant selon leurs désirs que selon leurs jugements évaluatifs, ce qui suggère que la voie directe entre cognition et motivation n'est pas suffisante pour générer des actions volontaires. Les croyances seraient donc incapables de nous motiver en l'absence de désirs.

Finalement, Schroeder et ses collègues notent qu'il n'existe qu'une circonstance connue dans laquelle les aires impliquées dans la cognition peuvent directement motiver des actions : le syndrome Gilles de la Tourette. Les patients atteints de ce syndrome sont victimes de tics difficilement contrôlables (notamment des hurlements), qui sont apparemment dus à une mauvaise inhibition de certains des signaux que les systèmes cognitifs envoient aux ganglions de la base. En clair, les croyances sont rarement suffisantes pour motiver des actions en l'absence de désirs, et les rares fois où elles le sont, les comportements en question sont des tics et des actes automatiques sur lesquels le sujet n'a aucun contrôle.

²⁹ Antti Kauppinen, « Moral internalism and the brain », *Social Theory and Practice*, 34(1), 2008, p. 1-24.

³⁰ Timothy Schroeder, Adina Roskies et Shaun Nichols, « Moral motivation », in J. Doris & The Moral Psychology Research Group (dir.), *The moral psychology handbook*, New York, Oxford University Press, 2010, p. 72-110.

Schroeder et ses collègues en concluent que les jugements moraux, comme les autres croyances, ne sont pas intrinsèquement motivants.

L'une des limites de cet argument est qu'il présuppose que les jugements moraux sont des croyances et, par là, la vérité du cognitivisme. Les données récoltées par Schroeder et ses collègues sont pertinentes quant à la vérité de la théorie huméenne de la motivation. Et cette théorie entraîne l'externalisme une fois combinée au cognitivisme, comme nous l'avons vu en introduction – plus précisément, nous avons vu que l'internalisme combiné à la théorie huméenne entraînait le non-cognitivisme, ce qui revient toutefois au même. Mais cet argument en faveur de l'externalisme est non seulement fragile (dans la mesure où le cognitivisme fait lui-même l'objet d'une controverse), mais aussi inutile pour les métaéthiciens, qui s'intéressent à l'internalisme principalement dans l'idée d'en tirer des leçons concernant la nature des jugements moraux.

a.iii. **La psychologie du jugement moral au secours de l'internalisme**

Jusqu'ici, nous n'avons rencontré que des usages externalistes des données empiriques. C'est qu'il est plus facile de rejeter l'internalisme que de le défendre : alors qu'un exemple de jugement moral non motivant suffit à le réfuter, on ne saurait l'établir au moyen d'un exemple de comportement motivé par un jugement moral. Par chance, comme nous l'avons vu dans la section précédente, les sciences empiriques ont d'autres choses à offrir que des contre-exemples : il est possible de mobiliser des théories systématiques afin de discuter de l'internalisme. C'est ce que font Schroeder et ses collègues, qui emploient les théories scientifiques à le rejeter. Mais, comme nous l'avons constaté, leur critique suppose que les jugements moraux sont réalisés uniquement au niveau des aires liées à la cognition supérieure. Or, certains théoriciens mettent en doute cette idée, en s'appuyant sur les théories sentimentalistes du jugement moral.

Revenons donc à nos psychopathes. Nous avons vu ci-dessus que ces sujets souffraient d'un déficit de maîtrise des concepts moraux. Ce déficit n'est pas à mettre sur le compte d'une moindre intelligence, les psychopathes n'étant pas moins intelligents que le reste de la population. Ils manifestent en revanche certains déficits émotionnels : ils ne ressentent ni culpabilité ni empathie. Certains théoriciens en concluent que leur incapacité à user des concepts moraux a à voir avec ces déficits. On a par ailleurs constaté que les jugements moraux des sociopathes différaient sensiblement de la moyenne. Confrontés à des dilemmes moraux, par exemple, ils se montrent plus utilitaristes que les sujets normaux³¹. Ces différences aussi pourraient s'expliquer par leurs troubles émotionnels – leurs réactions émotionnelles sont faibles, voire inexistantes³².

Cette curieuse corrélation entre déficits d'ordre moral et déficits d'ordre affectif a conduit plusieurs philosophes à développer des conceptions *sentimentalistes* du jugement moral, selon lesquelles les émotions font partie intégrante des jugements

³¹ Michael Koenigs, Liane Young, Ralph Adolphs, Daniel Tranel, Fiery Cushman, Marc D. Hauser et Antonio Damasio, « Damage to the prefrontal cortex increases utilitarian moral judgements », *Nature*, 446(7138), 2007, p. 908-911.

³² Damasio, *L'erreur de Descartes*.

moraux. Ainsi, Jesse Prinz³³ soutient que les jugements moraux sont une forme de réaction émotionnelle, et Nichols³⁴ défend l'idée selon laquelle ils naissent de la combinaison d'une règle et d'une réaction émotionnelle. Or, la plupart des théories contemporaines des émotions s'accordent pour dire que les réactions émotionnelles ont toujours pour composantes des « tendances à l'action » : nous tendons, par exemple, à fuir quand nous avons peur et à briser des objets sous le coup de la colère. En clair, les émotions seraient intrinsèquement motivantes³⁵. À supposer que les jugements moraux soient (ou impliquent nécessairement) des réactions émotionnelles, il faut bien en conclure qu'ils sont nécessairement motivants³⁶.

Bien que puissant à première vue, cet argument en faveur de l'internalisme présente indéniablement quelques faiblesses. Pour commencer, il repose sur une certaine conception du jugement moral, selon laquelle les jugements moraux ne peuvent être réduits à de simples croyances. Or, comme nous l'avons déjà mentionné, le principal intérêt de la thèse internaliste consiste précisément dans le fait qu'elle apporte un argument en faveur du non-cognitivism – la conception selon laquelle les jugements moraux ne sont pas de simples croyances. Comme l'argument de la section précédente, cet argument a ainsi le défaut de supposer résolue la controverse même à laquelle la thèse internaliste est censée mettre un terme. Il ne présente donc que peu d'intérêt pour les métaéthiciens.

Son second défaut est qu'il repose sur une certaine théorie des émotions, selon laquelle les tendances à l'action et l'aspect motivationnel d'une émotion font partie intégrante des émotions. Tout le monde s'accorde certes pour dire que les émotions nous motivent et possèdent certaines tendances à l'action. Mais cette thèse est susceptible d'une interprétation faible (les émotions tendent généralement à nous motiver) et d'une interprétation forte (les émotions suffisent toujours à nous motiver). Le présent argument présuppose l'interprétation forte : un internalisme au sujet des émotions selon lequel les émotions sont nécessairement motivantes. Or, il nous arrive de ressentir des émotions sans pour autant être motivés à agir. Comme mentionné dans la section précédente, certains problèmes neurologiques (maladies dégénératives ou lésions) entraînent des troubles de la motivation dont les victimes ont beaucoup de mal à agir. Ces troubles de la motivation s'accompagnent parfois de déficits émotionnels, comme dans la maladie de Parkinson, mais ils peuvent aussi apparaître en l'absence de tels déficits. Ainsi, certains patients souffrant de lésions bilatérales des ganglions de la base ont des troubles de la motivation. Ils semblent néanmoins conserver des réactions émotionnelles intactes (Schmidt et al., 2008). Ceci suggère que les émotions peuvent subsister en l'absence de motivations. À moins de soutenir que les affects de ces patients ne sont pas d'authentiques émotions, précisément parce qu'ils ne

³³ Jesse Prinz, *Gut reactions: a perceptual theory of emotion*, Oxford, Oxford University Press, 2004.

³⁴ Shaun Nichols, *Sentimental rules: on the natural foundations of moral judgment*, Oxford, Oxford University Press, 2004.

³⁵ Agnes Moors, « Theories of emotion causation: a review », *Cognition and Emotion*, 23(4), 2009, p. 625-662.

³⁶ Jesse Prinz, « The emotional basis of moral judgments », *Philosophical Explorations*, 9(1), 2006, p. 29-43.

sont pas motivants, il est tentant de nier l'existence d'un lien nécessaire entre émotions et motivation.

La solidité du présent argument dépend donc de celle de ses présupposés, qui demeurent assez controversés. Notons que des théories prometteuses ont récemment émergé, qui distinguent divers types de jugements moraux, dont certains seulement seraient ou impliqueraient des réactions affectives. Ces théories ont pour conséquence que l'internalisme est vrai pour certains types de jugements moraux, mais faux pour d'autres³⁷.

b. L'approche de la philosophie expérimentale

L'internalisme conceptuel étant une thèse a priori, il peut être évalué à la lumière de nos intuitions sur l'application du concept de jugement moral. Or, ce concept n'est pas technique : il relève en premier lieu de la psychologie ordinaire. On aurait donc tort de confronter l'internalisme aux intuitions des seuls philosophes ; celles des locuteurs ordinaires ne sont pas moins pertinentes. Il est même vraisemblable qu'elles le soient davantage, dans la mesure où elles sont à l'abri des biais théoriques auxquels sont exposés les philosophes³⁸.

Forts de ces considérations, les philosophes expérimentaux entreprennent de sonder empiriquement nos intuitions au sujet d'expériences de pensée philosophiques. Leur approche combine la pratique de l'analyse conceptuelle avec des méthodes expérimentales qui étaient jusqu'alors l'apanage des sciences cognitives³⁹. Dans le cas qui nous occupe, les philosophes expérimentaux ont recours à cette même approche afin de déterminer si les intuitions des non-philosophes confirment ou réfutent l'internalisme.

L'une des premières études allant dans ce sens est relatée dans un article de Nichols⁴⁰. Ce dernier y rapporte avoir soumis ses participants à l'expérience de pensée suivante :

John est un criminel psychopathe. Bien qu'il soit d'âge adulte et d'intelligence normale, blesser autrui ne provoque chez lui aucune réaction émotionnelle. John a blessé et parfois même tué des gens dans le simple but de leur voler leur argent. De son propre aveu, il sait que nuire aux autres est mal (*wrong*) mais se fiche de mal agir.

Question : John comprend-il vraiment qu'il est immoral de nuire à autrui ?

À cette question, près de 85% des participants ont répondu par l'affirmative, ce qui suggère à première vue qu'une large proportion de la population ne voit pas d'objection à attribuer un jugement moral à quelqu'un qui ne se conforme pas à ce jugement.

L'expérience de Nichols présente néanmoins plusieurs limites. Premièrement, il n'est pas clair que John n'ait aucune motivation à agir moralement : même s'il n'est pas *pro toto* motivé à agir conformément à ses jugements moraux, il se

³⁷ Christine Clavien, « Jugements moraux et motivation à la lumière des données empiriques », *Studia Philosophica*, 68, 2009, p. 179-206.

³⁸ Gunnar Björnsson, John Eriksson, Caj Strandberg, Ragnar Francén Olinder et Fredrik Björklund, « Motivational internalism and folk intuitions. », *Philosophical Psychology*, 28(5), 2015, p. 715-734.

³⁹ Florian Cova, Julien Dutant, Edouard Machery, Joshua Knobe, Shaun Nichols et Nahmias, Eddy, *La Philosophie Expérimentale*. Paris, Vuibert, 2012.

⁴⁰ Nichols, « How psychopaths threaten moral rationalism ».

pourrait qu'il le soit *pro tanto*⁴¹. Deuxièmement, la question soumise aux participants porte sur la compréhension plutôt que sur le jugement. Or, il se pourrait que John soit incapable de juger qu'il est immoral de nuire à autrui même s'il comprend qu'il est immoral de nuire à autrui.

Fort heureusement, l'investigation empirique des intuitions au sujet de la motivation morale a connu de récents développements, et des scénarios plus précis ont été utilisés. Caj Strandberg et Fredrik Björklund⁴² ont ainsi eu recours au scénario suivant :

Anne regarde un reportage consacré à la famine au Soudan dans lequel on voit des gens qui meurent de faim et cherchent désespérément de quoi se nourrir. Anne ne souffre pas de troubles mentaux et fonctionne normalement : elle n'est ni dépressive, ni apathique, ni émotionnellement perturbée, ni psychopathe, ni quoi que ce soit de ce genre. Alors qu'elle regarde le reportage, Anne n'est pas motivée du tout, pas le moins du monde, à donner de l'argent à ceux qui sont en train de mourir de faim.

Question : Se peut-il qu'Anne pense qu'il est moralement requis de donner une part de son argent à ceux qui meurent de faim, quand bien même elle n'est pas le moins du monde motivée à le faire ?

À cette question, 79% des participants ont répondu par l'affirmative : il est possible qu'Anne juge qu'elle a l'obligation de consacrer une part de sa fortune à lutter contre la famine sans être le moins du monde disposée à le faire. Puisque le scénario précise qu'Anne n'est ni dépressive, ni apathique, ni émotionnellement perturbée, ni psychopathe, ces intuitions semblent aller directement à l'encontre non seulement de l'internalisme simple, mais aussi de l'internalisme conditionnel. Un autre scénario a été développé afin de tester l'internalisme collectif, qui décrit une société peuplée d'individus tout aussi indifférents qu'Anne au sort des victimes de la famine. À la lecture de ce scénario, 72% des participants ont considéré qu'il était possible que ces individus jugent réellement qu'il est requis de consacrer une part de leur argent à la lutte contre la famine.

Les non-philosophes paraissent donc capables de concevoir des cas de jugements moraux non motivants, ce qui pose problème pour l'internalisme conceptuel. De fait, ces résultats empiriques semblent remettre en cause l'idée selon laquelle le concept même de jugement moral impliquerait que ces jugements sont motivants, et par là l'internalisme.

Il n'est toutefois pas clair qu'on tienne là un argument définitif contre l'internalisme. Les investigations empiriques consacrées à l'internalisme n'en sont qu'à leurs balbutiements et doivent encore surmonter certaines difficultés méthodologiques⁴³. Il se pourrait donc que de nouvelles études viennent mettre en doute ces résultats et nous incitent à accepter une conclusion opposée. C'est ce qui s'est produit, par exemple, avec les études consacrées au libre-arbitre, où le

41 Il est vrai que la vignette présentée ici précise qu'il « s'en fiche ». Cependant, un problème récurrent de ce genre d'étude est que les gens ont tendance à ne pas toujours croire ce que disent les personnages décrits dans les vignettes, et à projeter sur eux leurs propres valeurs morales.

42 Caj Strandberg et Fredrik Björklund, « Is moral internalism supported by folk intuitions? », *Philosophical Psychology*, 26(3), 2013, p. 319-335.

43 Björnsson *et al.*, « Motivational internalism and folk intuitions ».

consensus s'est graduellement déplacé de la thèse selon laquelle les gens croient le libre-arbitre et le déterminisme incompatibles vers l'idée selon laquelle ils les jugent compatibles dès lors qu'ils maîtrisent ces deux concepts⁴⁴.

Une deuxième limite de cet argument est qu'il présuppose que les intuitions des non-philosophes sont pertinentes pour le débat philosophique. Or, il se pourrait que les résultats mêmes de ces études viennent un jour menacer ce présupposé méthodologique, en montrant que les intuitions des non-philosophes sont instables et biaisées, et donc insuffisamment fiables pour servir de données dans le débat philosophique. Il semble, par exemple, que les gens soient plus susceptibles d'avoir des intuitions internalistes quand les jugements concernés portent sur des actions interdites ou obligatoires, plutôt que sur des actions optionnelles et surrogatoires. Autrement dit, le contenu de nos croyances morales influencerait nos intuitions au sujet de l'internalisme⁴⁵.

5. Conclusion

Le débat concernant la force motivationnelle des jugements moraux demeure ouvert : ni l'internalisme ni l'externalisme n'a jusqu'ici été établi. Il est donc difficile de conclure ce survol du débat par un verdict définitif. Nous nous contenterons par conséquent de résumer le chemin parcouru.

Dans un premier temps, nous avons constaté que le débat philosophique n'était pas concluant. Contrairement aux allégations de ses opposants, l'externalisme est en mesure de rendre compte de notre motivation à agir conformément à nos jugements moraux, y compris lorsque cette motivation survit à une conversion morale. En outre, les intuitions des philosophes s'alignant la plupart du temps sur les théories auxquelles ils souscrivent, les expériences de pensée impliquant des sujets amoraux ne permettent pas non plus de trancher le débat. Il va de soi pour les externalistes que ces individus forment de véritables jugements moraux, tandis que les internalistes considèrent qu'ils font un usage entre guillemets des termes moraux.

D'où l'idée de recourir aux données expérimentales. Mais force est de constater que l'apport de la psychologie et des neurosciences au débat sur l'internalisme a jusqu'ici été limité. Premièrement, quand les externalistes soutiennent que les psychopathes et les sociopathes constituent des contre-exemples à l'idée que les jugements moraux sont nécessairement motivants, les internalistes nient que ces individus fassent d'authentiques jugements moraux. Que les amoralistes soient réels ou imaginaires, le problème demeure : il se pourrait qu'ils emploient en fait le langage moral pour décrire les attitudes de leur communauté. Deuxièmement, les externalistes qui s'appuient sur les données neuroscientifiques présupposent la vérité du cognitivisme, ce qui est assez ennuyeux, le principal enjeu du débat étant précisément d'élucider la nature des jugements moraux. Enfin, on peut reprocher une erreur symétrique aux philosophes qui mobilisent les données de la psychologie du jugement moral afin de défendre l'internalisme : ils présupposent

⁴⁴ Florian Cova et Yasuko Kitano, « Experimental philosophy and the compatibility of free will and determinism: a survey », *Annals of the Japan Association for Philosophy of Science*, 22, 2014, p. 17-37.

⁴⁵ Derek Leben et Kristine Wilckens, « Pushing the intuitions behind moral internalism », *Philosophical Psychology*, 28(4), 2015, p. 510-528.

la vérité du non-cognitivism. Il est clair que si les jugements moraux sont des états conatifs, ils sont nécessairement motivationnels. Mais il est dialectiquement étrange de défendre l'internalisme sur la base du non-cognitivism.

Les arguments qui nous semblent les plus intéressants émanent de la philosophie expérimentale. Pour l'heure, il semble que ces arguments plaident plutôt en faveur de l'externalisme : la plupart des sujets interrogés attribuent volontiers des jugements moraux aux amoralistes. Ce pan de la recherche n'en est évidemment qu'à ses débuts. Mais si les données dont on dispose actuellement sont à manipuler avec précaution, il n'est pas déraisonnable d'espérer que cette approche mette un terme à la controverse sur l'internalisme.

Références

- Bedke, Matthew, « Moral judgments purposivism : saving internalism from amoralism », *Philosophical Studies*, 144, 2009, p. 189-209.
- Björnsson, Gunnar, John Eriksson, Caj Strandberg, Ragnar Francén Olinder et Fredrik Björklund, « Motivational internalism and folk intuitions », *Philosophical Psychology*, 28(5), 2015, p. 715-734.
- Blair, Robert James, « A cognitive developmental approach to morality: investigating the psychopath », *Cognition*, 57, 1995, p. 1-29.
- Blackburn, Simon, *Ruling passions : a theory of practical reasoning*, New York, Oxford University Press, 1998.
- Brink, David Owen, « Externalist moral realism », *Southern Journal of Philosophy*, supplement, 1986, p. 23-42.
- *Moral realism and the foundations of ethics*, Cambridge, Cambridge University Press, 1989.
- Cholbi, Michael, « Belief attribution and the falsification of motive internalism », *Philosophical Psychology*, 19(5), 2006, p. 607-616.
- Cima, Maaïke, Franca Tonnaer et Marc D. Hauser, « Psychopaths know right from wrong but don't care », *Social Cognitive and Affective Neuroscience*, 5(1), 2010, p. 59-67.
- Clavien, Christine, « Jugements moraux et motivation à la lumière des données empiriques », *Studia Philosophica*, 68, 2009, p. 179-206.
- Copp, David, « Belief, reason, and motivation: Michael Smith's "The moral problem" », *Ethics*, 108(1), 1997, p. 33-54.
- Cova, Florian, Julien Dutant, Edouard Machery, Joshua Knobe, Shaun Nichols et Nahmias, Eddy, *La Philosophie Expérimentale*. Paris, Vuibert, 2012.
- Cova, Florian et Yasuko Kitano, « Experimental philosophy and the compatibility of free will and determinism: a survey », *Annals of the Japan Association for Philosophy of Science*, 22, 2014, p. 17-37.
- Damasio, Antonio, *L'erreur de Descartes*, Paris, Odile Jacob, 2006.
- Haidt, Jonathan, « The emotional dog and its rational tail: a social intuitionist approach to moral judgment », *Psychological review*, 108(4), 2001, p. 814-834.
- Hare, Richard Mervyn, *The language of morals*, Oxford, Oxford University Press, 1952
- Joyce, Richard, *The myth of morality*, Cambridge, Cambridge University Press, 2001.
- Kauppinen, Antti, « Moral internalism and the brain », *Social Theory and Practice*, 34(1), 2008, p. 1-24.

- Koenigs, Michael, Liane Young, Ralph Adolphs, Daniel Tranel, Fiery Cushman, Marc D. Hauser et Antonio Damasio, « Damage to the prefrontal cortex increases utilitarian moral judgements », *Nature*, 446(7138), 2007, p. 908-911.
- Leben, Derek et Kristine Wilckens, « Pushing the intuitions behind moral internalism », *Philosophical Psychology*, 28(4), 2015, p. 510-528.
- Levy, Neil, « The responsibility of the psychopath revisited », *Philosophy, Psychiatry, & Psychology*, 14(2), 2007, p. 129-138.
- Lillehammer, Hallvard, « Smith on moral fetishism », *Analysis*, 57(3), 1997, p. 187-195.
- McNaughton, David, *Moral vision: an introduction to ethics*, Oxford, Blackwell, 1988.
- Moors, Agnes, « Theories of emotion causation: a review », *Cognition and Emotion*, 23(4), 2009, p. 625-662.
- Nichols, Shaun, « How psychopaths threaten moral rationalism », *The Monist*, 85(2), 2002, p. 285-303.
- *Sentimental rules: on the natural foundations of moral judgment*, Oxford, Oxford University Press, 2004.
- Platon, *La République*, Paris, Garnier Flammarion, 2002.
- Prinz, Jesse, *Gut reactions: a perceptual theory of emotion*, Oxford, Oxford University Press, 2004.
- « The emotional basis of moral judgments », *Philosophical Explorations*, 9(1), 2006, p. 29-43.
- Railton, Peter, « Moral realism », *Philosophical Review*, 95, 1986, p. 163-207.
- Roskies, Adina, « Are ethical judgments intrinsically motivational? Lessons from “acquired sociopathy” », *Philosophical Psychology*, 16(1), 2003, p. 51-66.
- « Patients with ventromedial frontal damage have moral beliefs », *Philosophical Psychology*, 19(5), 2006, p. 617-627.
- Schmidt, Liane, Baudoin Forgeot d'Arc, Gilles Lafargue, Damien Galanaud, Virginie Czernecki, David Grabli, Michael Schüpbach, Andreas Hartmann, Richard Levy, Bruno Dubois et Mathias Pessiglione, « Disconnecting force from money: effects of basal ganglia damage on incentive motivation », *Brain*, 131(5), 2008, p. 1303-1310.
- Schroeder, Timothy, Adina Roskies et Shaun Nichols, « Moral motivation », in J. Doris & The Moral Psychology Research Group (dir.), *The moral psychology handbook*, New York, Oxford University Press, 2010, p. 72-110.
- Smith, Michael, *The Moral Problem*. Oxford: Wiley-Blackwell, 1994.
- Strandberg, Caj et Fredrik Björklund, « Is moral internalism supported by folk intuitions? », *Philosophical Psychology*, 26(3), 2013, p. 319-335.
- Sigrun Svavarsdottir, « Moral cognitivism and motivation », *The Philosophical Review*, 108(2), 1999, p. 161-219.
- Williams, Bernard, *Moral luck*, Cambridge, Cambridge University Press, 1981.