

Swiss Philosophical Preprint Series

37

François Jaquet

Cognitivisme, Beurk! Non-Cognitivisme, Hurrah!

Added 20/02/2009

ISSN 1662-937X

© François Jaquet

COGNITIVISME, BEURK ! NON-COGNITIVISME, HURRAH !

La philosophie morale est traditionnellement divisée en trois sous-disciplines : l'éthique appliquée, qui, comme son nom l'indique, s'intéresse aux positions à adopter sur des sujets pratiques ; l'éthique normative, où s'opposent un ensemble de théories sur ce qui devrait être, sur ce qui est bon/mauvais, etc. ; et la méta-éthique, qui étudie des questions non morales, mais relatives à la morale¹. Cette dernière définition peut paraître floue, mais c'est un flou que rend nécessaire l'hétérogénéité propre à la discipline. En effet, celle-ci est elle-même divisée en plusieurs domaines : l'épistémologie morale, l'ontologie morale, la psychologie morale, la phénoménologie morale et la logique morale (l'analyse des concepts moraux). Il n'est pas impossible de croiser les domaines, à l'instar de David McNaughton qui oppose une théorie ontologique – le réalisme – à une théorie épistémologique – le non-cognitivism. Le réalisme moral affirme l'existence de propriétés morales non réductibles à des propriétés non morales : un acte est condamnable si et seulement s'il possède, à côté de ses propriétés non morales (il s'agit d'un acte délibéré, prémédité, etc.), la propriété d'être condamnable. Et, quant à lui, le cognitivism soutient qu'une opinion morale n'est rien de plus qu'un état cognitif, qu'une croyance. Ce que nie le non-cognitivism. Ainsi, selon McNaughton, le principal argument contre le réalisme moral est non cognitiviste. Le voici :

- (P1) Si la théorie humienne de la motivation et l'internalisme moral sont vrais, alors le cognitivism moral est faux.
- (P2) (Or) la théorie humienne de la motivation est vraie,
- (P3) (Et) l'internalisme moral est vrai,
- (P4) (Et) si le cognitivism moral est faux, alors le réalisme moral est faux.
- (C) (Donc) le réalisme moral est faux.

L'argument étant clairement valide, notre travail sera divisé en quatre parties, dont chacune étudiera la vérité d'une de ses prémisses. Si tout se passe bien, nous devrions pouvoir conclure par le rejet du réalisme. Ou pas ; nous verrons.

¹ Il semble qu'on ne peut pas dire que la méta-éthique s'occupe de questions ou de propositions de second ordre (même s'il n'est pas interdit de le regretter). Par exemple, la question d'ontologie morale « Les propriétés morales existent-elles ? » n'est une question à propos d'aucune question. Et « Elles existent », une proposition à propos d'aucune proposition.

1. Logique du trilemme (P1)

La théorie humienne de la motivation dit que, nécessairement, si x est motivé à agir, alors x possède une croyance moyens-fin et un désir. Une croyance ne suffit pas, et un désir non plus. Ainsi, pour que je me mette sérieusement à la rédaction de ce travail, il a fallu non seulement que je croie que le sérieux de mon travail est une condition de l'obtention de mes crédits, mais aussi que je désire les obtenir. En l'absence d'un tel désir, il est probable – sauf à avoir du temps libre et un goût rare pour la rédaction – que je n'aurais pas rédigé. De même, si j'avais pensé que je pouvais obtenir mes crédits sans travailler sérieusement (mais dans le cadre de la déontologie académique, bien sûr), il est probable que j'aurais été moins rigoureux.

L'internalisme est une position en psychologie morale, selon laquelle, nécessairement, si x a une opinion morale, alors x est motivé à agir conformément (y compris à s'abstenir dans le cas où le jugement serait négatif). Admettons que j'estime avoir une obligation morale à (faire le nécessaire pour) obtenir mes crédits. C'est un devoir envers moi-même, et je ne suis pas minimaliste moral pour un sou. Admettons. Alors, si l'internalisme moral est vrai, je suis nécessairement motivé à obtenir mes crédits.

Notre première prémisse, (P1), affirme donc que si l'opinion morale requiert la motivation morale, et que celle-ci requiert à son tour un désir, alors l'opinion morale requiert un désir. Pas de jugement moral sans désir d'agir conformément. Elle est logiquement vraie².

2. Quand on peut, on veut³ (P2)

Michael Smith (1987 : 55) propose l'argument suivant en faveur de la théorie humienne de la motivation :

(Q1) Avoir une raison motivante, *c'est, inter alia*, avoir un but.

(Q2) (Or) avoir un but, *c'est* être dans un état auquel le monde doit s'adapter.

(Q3) (Et) être dans un état auquel le monde doit s'adapter, *c'est* désirer.

(C2) (Donc) avoir une raison motivante, *c'est, inter alia*, désirer,

² Car elle a la forme « $((P \rightarrow Q) \wedge (Q \rightarrow R)) \rightarrow (P \rightarrow R)$ », dont nous pourrions prouver la vérité logique si c'était nécessaire. Ce qui n'est manifestement pas le cas.

³ Le bon mot nécessite, paraît-il, une explication. Que voici : la thèse humienne est qu'il n'y a pas de motivation sans désir. Il n'y a donc pas d'action sans désir, puisqu'il ne saurait de toute évidence y avoir d'action sans motivation. En conséquence, la thèse humienne implique qu'il n'y a pas d'action sans désir. Si on ne veut pas agir, on ne peut pas agir. Par contraposition, quand on peut, on veut.

autrement dit, pas de motivation sans désir. Les deux premières prémisses sont conceptuelles, analytiques. Sauf à le contester (et puisque l'argument est manifestement valide), c'est donc à la troisième que le critique devra s'attaquer. Et c'est ce que fait McNaughton.

L'expression « un état auquel le monde doit s'adapter » réfère à une distinction opérée par Smith (1987 : 51) entre deux « directions d'adaptation » (*directions of fit*) qui oppose, selon lui, les états cognitifs aux états non cognitifs : les premiers doivent s'adapter au monde pour être satisfaits – disons qu'ils sont passifs –, alors que c'est le monde qui doit s'adapter aux seconds pour que ceux-ci soient satisfaits – appelons-les « actifs ».

L'anti-huméen peut soit rejeter cette distinction, soit distinguer entre les croyances motivationnelles (qui ont les deux directions d'adaptation) – par exemple, connaître les exigences morales – et les croyances non motivationnelles. Si les croyances motivationnelles existent, alors nul besoin de faire appel à des états mentaux non cognitifs – qui n'ont que la direction d'adaptation le-monde-doit-s'adapter-à-moi – pour expliquer l'action. La troisième prémisses, (Q3), est fautive, car les croyances motivationnelles à la fois sont telles que le monde doit s'y adapter et sont des états cognitifs, ce dont elle nie la possibilité.

La première option étant peu attrayante, McNaughton opte pour la seconde. En faveur de l'existence de croyances motivationnelles, il avance deux arguments.

Les croyances motivationnelles existent (I)

Si les états mentaux étaient soit passifs soit actifs, mais jamais les deux à la fois, – et si, donc, les croyances motivationnelles n'existaient pas –, nous distinguerions l'élément cognitif de l'élément affectif dans nos évaluations, et en particulier dans nos évaluations morales. Mais ce n'est pas le cas : quand nous évaluons moralement, nous ne le faisons pas en deux temps – un premier, cognitif, pour appréhender les faits, et un second, affectif, pour les évaluer. Au contraire, nous percevons le caractère moralement bon ou mauvais des actions de la même manière que nous percevons leurs autres propriétés. C'est donc que les croyances motivationnelles existent. C'est en tout cas l'argument de McNaughton.

Mais la première prémisses de cet argument ne résiste pas à un examen critique : il est faux que la présence d'éléments cognitifs et non cognitifs dans un acte d'évaluation nécessite deux phénoménologies distinctes – phénoménologie de la croyance et phénoménologie du désir. Et pour cause : les désirs ne sont pas essentiellement des états mentaux phénoménologiques. Comme le dit Smith,

[a] father may actually feel the prick of [the desire that his children do well] from time to time; in moments of reflection on their vulnerability, say. But such occasions are not the norm. Yet we wouldn't ordinarily think that he loses this desire during those periods when he lacks such feelings. (1987 : 48)

Il est tout à fait possible de désirer inconsciemment la réalisation d'un état de fait. Si McNaughton préfère manifestement l'approche phénoménologique du désir, il n'explique pas pourquoi il faudrait rejeter les arguments que Smith lui oppose.

Les croyances motivationnelles existent (II)

La deuxième manœuvre de McNaughton consiste à ranger les désirs dans la catégorie des états cognitifs. Si les désirs sont des états cognitifs, alors même s'ils sont nécessaires à la motivation, la thèse humienne tombe, car elle soutient que des états *non cognitifs* sont nécessaires à une telle explication. Cette tactique de McNaughton est bizarre car elle revient à affirmer que tous les désirs sont des croyances motivationnelles (puisque tous les états cognitifs ont la direction je-dois-m'adapter-au-monde, et que ceux qui ont aussi la direction le-monde-doit-s'adapter-à-moi sont des croyances motivationnelles), un point terminologique qui ne devrait pas faire avancer le débat. Mais voyons son argument.

Pour comprendre les désirs il est nécessaire et suffisant d'adopter une perception de la chose désirée. Or si la perception est nécessaire et suffisante à la compréhension des désirs, il faut conclure que ceux-ci sont des états cognitifs. Donc les désirs sont des états cognitifs. Cet argument soulève deux objections.

Premièrement, pour expliquer les désirs, le non-cognitivist et le cognitiviste peuvent tous deux faire appel à la perception des propriétés ordinaires des actions. Admettons, pour les besoins de l'argument, que le cognitiviste peut faire appel, en plus, à la propriété de désirabilité de l'action. Le problème est le suivant : en quoi dire « l'action *A* a la conjonction de propriétés descriptives *F* plus la propriété évaluative *G* » serait une meilleure explication que « *A* a la conjonction de propriétés descriptives *F* et je désire que *F* soient instanciée ? » Soyons sérieux ! Soit la perception de *G* dépend de la perception de *F*, soit non. Si elle dépend de la perception de *F*, alors l'explication ne sera pas meilleure que de dire qu'un état non cognitif dépend de la perception de *F*. Par contre, si elle en est indépendante, on voit mal comment l'explication en est une : comment faire percevoir une propriété telle que la désirabilité d'un objet à quelqu'un qui ne le désire pas, sinon en l'aidant à percevoir ses propriétés descriptives ? Comment savoir qu'un objet est désirable sans connaître ses

propriétés ordinaires ? Il semble que, dans tous les cas, l'existence d'une propriété de désirabilité n'aide en rien à l'explication de la motivation. La première prémisse de l'argument de McNaughton est simplement fausse.

Deuxièmement, personne ne contestera que des états cognitifs puissent causer des états non cognitifs. Mais ce qui est en jeu, c'est l'identité. Et McNaughton ne prouve pas que les désirs *sont* des états cognitifs ; seulement qu'ils sont *causés par* des états cognitifs. Peut-être que voir la douleur comme un masochiste (ses propriétés ordinaires) suffit à comprendre ou même partager son désir. Mais ça montre seulement que c'est une cause suffisante, non pas qu'en percevant les propriétés on a, entre autres, perçu des propriétés extraordinaires (désirabilité de la douleur physique).

Plus mystérieusement, selon McNaughton « [l]e réaliste prétend que la compréhension vient de l'appréciation, *même si pas nécessairement du partage*, de l'expérience de la personne dont il essaie d'expliquer l'action de l'intérieur » (112, nous traduisons et soulignons). Il distingue donc appréciation (*i.e.* perception) et partage (*i.e.* désir). Mais comment pourrait-on percevoir de la même manière sans avoir un désir correspondant si le désir relevait de la perception ? Si percevoir un objet *O* comme un individu *I* signifie, entre autres, percevoir les propriétés évaluatives de *O* que perçoit *I*, alors si *I* perçoit en *O* une propriété de désirabilité, toute personne qui perçoit *O* comme *I*, percevra nécessairement cette propriété. Enfin, soit percevoir cette propriété revient à désirer *O*, soit non. Dans le premier cas, la remarque de McNaughton est simplement fausse. Dans le second, la propriété n'est que contingemment motivationnelle, ce que peuvent accepter les non-cognitivistes – « on peut logiquement désirer n'importe quoi » –, mais pas les cognitivistes internalistes. Ceci montre bien que désirs et états cognitifs, s'ils peuvent entretenir des relations causales, éventuellement nécessaires, n'entretiennent nécessairement pas de relations d'identité. Et que McNaughton – s'il n'en tire pas les justes conclusions – le reconnaît. La deuxième prémisse doit être abandonnée.

Les croyances motivationnelles n'existent pas

Quoi qu'il faille penser de ces deux arguments de McNaughton en faveur de l'existence de croyances motivationnelles, Smith, ayant anticipé cette objection à la troisième prémisse de son argument, (Q3), lui avait apporté une réponse (1987 : 56) : un seul et même état mental ne peut avoir deux directions d'adaptation. Et voici pourquoi. Les états mentaux sont des états dispositionnels. (Une disposition est une tendance, actualisée ou non, à produire tel ou tel comportement dans telle ou telle circonstance. Admettons que les désirs sont des dispositions.

Alors désirer obtenir mes crédits, c'est avoir tendance à rédiger un beau travail si je crois que c'en est une condition nécessaire.) Le problème est que les états mentaux passifs manifestent des dispositions contraires à celles qui caractérisent les états mentaux actifs. La croyance que P aura tendance à disparaître avec la perception que $\neg P$, alors que le désir que P aura tendance à persister avec la perception que $\neg P$. Puisqu'un seul et même état mental ne peut avoir deux tendances contraires (disparaître et persister), un seul et même état mental ne peut être passif et actif, *i.e.* avoir les deux directions d'adaptation.

Si McNaughton n'a vraisemblablement pas été convaincu, il ne prend pas la peine d'expliquer pourquoi et se contente d'affirmer, avec un air d'originalité, qu'il est possible pour un état mental d'avoir les deux directions d'adaptation.

3. Faites ce que je dis, pas ce que je fais ? (P3)

L'internalisme est la thèse en psychologie morale, vieille sinon comme le monde au moins comme la philosophie morale, selon laquelle une opinion morale entraîne une motivation à agir conformément. Et que ce lien est nécessaire. Comme McNaughton, nous la trouvons intuitivement plausible : « [c]omment pourrions-nous soutenir qu'une action particulière était injuste et pourtant ne voir aucune raison de ne pas agir de cette manière ? » (1988 : 22, notre traduction)

C'est une donnée du problème que nous sommes motivés à agir par nos opinions morales. Et il nous semble que c'est en outre une condition nécessaire à l'intérêt de la philosophie morale. Un genre de postulat de la discipline. En effet, si la moralité des actions était indifférente aux choix que nous faisons, pourquoi chercher à la déterminer ? Internalistes et externalistes s'accordent donc sur ce point. (Bien sûr, il y a, comme nous le verrons, des contre-exemples – apparents, selon les internalistes.) Ce qui les oppose, c'est la nature du lien : pour les premiers, la relation entre opinion morale et motivation est nécessaire, alors que pour les seconds, elle est simplement empirique, contingente. Et l'issue du débat pourrait bien se jouer par inférence à la meilleure explication de la tendance à guider l'action qu'ont les opinions morales. Pour les externalistes, les opinions morales étant des croyances – les externalistes sont rarement non-cognitivistes –, elles réfèrent à des propriétés des objets, et ces propriétés sont attractives ou répulsives, désirables ou indésirables et, par là même, motivantes (pour l'action ou l'abstention) et c'est un fait contingent qu'elles sont souvent motivantes dans le sens conforme à l'évaluation. Pour les internalistes, les opinions morales sont des désirs (cognitifs ou non), et les désirs sont intrinsèquement motivants.

Reste que si la relation est contingente, expliquer une telle régularité est (sauf peut-être pour un partisan de l'égoïsme moral) difficilement explicable. Bref, la balle est dans le camp des externalistes. Qui ne tardent pas à s'en débarrasser, avec deux problèmes pour l'internalisme : la faiblesse de la volonté et l'amoralisme.

Oui, je devrais m'y mettre sérieusement, mais...

On dit de la volonté d'un individu qu'elle est faible lorsque, librement et intentionnellement, il ne fait pas ce qu'il juge qu'il devrait faire, la meilleure action (une abstention, le cas échéant), tout bien considéré. La faiblesse morale est une sous-catégorie de la faiblesse de la volonté : elle se produit lorsqu'un individu, librement et intentionnellement, ne fait pas ce qu'il juge être son devoir *moral*. On pourrait penser que la faiblesse morale est logiquement impossible si l'internalisme est vrai. En effet, dans ce cas, si x dit qu'il est de son devoir de *Fer*, alors, nécessairement, x désire *Fer*, et si x désire *Fer*, alors x est motivé à *Fer* (même si un désir n'était pas nécessaire à la motivation, il lui serait quand même suffisant, conjointement à une croyance). Et par contraposition, si x n'est pas motivé à *Fer*, alors x ne désire pas *Fer*, et, en vertu de l'internalisme, x ne peut penser qu'il est de son devoir de *Fer*. Pourtant si la faiblesse de la volonté est possible, ces deux chaînes de conditionnelles sont fausses. Et la faiblesse morale nous semble non seulement une possibilité logique, mais une réalité quotidienne (ou courante, si l'on veut être optimiste). C'est ce qu'on pourrait penser, mais on aurait tort.

En principe, il suffit à l'internaliste de fournir une explication à la faiblesse morale. C'est-à-dire, une explication qui ne fonctionne pas nécessairement pour la faiblesse de la volonté. Ainsi, McNaughton nous explique (1988 :126-7) que la motivation à agir moralement peut ne pas être déterminante, dans le cas, par exemple, où d'autres désirs entrent en contradiction avec elle. (Jules aimerait devenir végétalien. Mais il aime tellement saupoudrer ses pâtes de fromage râpé...) Cette explication fonctionne à condition que par « x est motivé à *Fer* » on n'entende pas « x est motivé à *Fer*, tout bien considéré ». Si c'était le cas, aucun autre désir ne pourrait venir contrebalancer la motivation à agir moralement. Plus précisément, on pourrait distinguer entre deux formes – *prima facie* et absolue – de motivation, et souligner que seule la seconde implique, conjointement à la thèse internaliste, l'impossibilité de la faiblesse morale. En effet, si une opinion morale implique la motivation, tout bien considéré, à agir conformément, alors elle implique aussi l'action.

Mais de cette distinction en découle une autre, entre deux formes correspondantes d'internalisme, que Donald Davidson appelle respectivement « extrême » et « atténuée »

(2008 : 43)⁴. Selon l'internalisme extrême, une opinion morale implique une motivation, tout bien considéré, absolue, à agir conformément. Et donc une action conforme. La faiblesse de la volonté est impossible. Au contraire, selon l'internalisme modéré, une opinion morale implique une motivation *prima facie*, laquelle, par définition, n'exclut pas l'existence de motivations contraires. Ces autres motivations pouvant, le cas échéant, dépasser en force la motivation morale, la faiblesse morale est possible.

Une forme d'internalisme peut donc être sauvée. Mais nous sommes en présence d'une nouvelle distinction. Nous avons besoin, pour notre argument antiréaliste, que « l'internalisme » soit vrai. Or si la faiblesse morale existe, on ne peut pas dire que l'internalisme est vrai ; il est vrai sous l'une de ses formes – la plus modérée –, faux sous l'autre. Il nous faut donc nous assurer que la première prémisse de notre argument antiréaliste, (P1), est vraie même si nous y remplaçons « internalisme » par « internalisme modéré ».

(P1) affirme maintenant que si l'opinion morale requiert la motivation morale *prima facie*, et que toute motivation requiert à son tour un désir, alors l'opinion morale requiert un désir. Et, par chance, elle est toujours logiquement vraie (puisque une motivation *prima facie* est une motivation quand même). La qualité de l'argument n'est pas compromise.⁵

Les critiques ont objecté à Davidson que si un individu considère que l'action A est la meilleure *tout court*, cela revient à dire qu'elle est la meilleure tout bien considéré, ce qui exclurait les motivations contraires. Mais ceci ne doit pas nous inquiéter puisque, en ce qui concerne l'internalisme moral, l'individu en question doit considérer que A est la meilleure option *moralement parlant* (vs tout court). Rien (sinon la morale) ne lui interdit de prendre en considération d'autres paramètres.⁶

Amoraliste, méchant, hypocrite et akratique

Un amoraliste est quelqu'un qui n'attache pas d'importance à la moralité de ses actions en général, ou d'une de ses actions en particulier – s'il s'agit d'un amoraliste occasionnel, disons.

⁴ Davidson distingue entre les jugements – et non les motivations – *prima facie* et absolus. Mais le résultat, s'il diffère dans le cas de la faiblesse de la volonté, est identique pour la faiblesse morale. Une opinion qui implique une motivation *prima facie* est une opinion *prima facie*, et les opinions morales peuvent être considérées comme des opinions *prima facie*.

⁵ La lecture de Furrer (2009 : 7) – excellent texte s'il en est – nous a permis de prendre conscience du fait que notre internalisme atténué n'est autre que l'externalisme de David Brink. Ce qui, au vu de nos conclusions, a la conséquence étonnante – n'est-ce pas ? – que l'externalisme brinkien est dans l'incapacité de rejeter notre argument anticognitivistique par sa troisième prémisse.

⁶ Voir note 3.

Au contraire, un méchant (au sens technique qui nous intéresse) est quelqu'un qui attache de l'importance à la moralité de ses actions. Elle est pour lui une raison de s'en abstenir.

De prime abord, ces deux positions semblent réfuter l'internalisme moral. Prenons l'amoralisme, pour commencer. Il implique apparemment la possibilité pour un individu d'agir alors qu'il juge son action immorale, ainsi que celle de ne pas agir alors qu'il juge moralement obligatoire de le faire. Or nous avons vu que l'internalisme nie chacune de ces possibilités. Pourtant, l'amoralisme, au moins ponctuel, nous est relativement familier.

McNaughton (1988 : 136-7) présente la solution non-cognitiviste standard au problème de l'amoralisme. Celui ou celle qui utilise le terme « mauvais » pour décrire sa propre action, sans être pour autant en état de faiblesse de la volonté, montre par là même qu'il ne maîtrise pas sa signification : il le traite comme un terme descriptif alors que c'est un terme évaluatif, l'utilise dans un sens différent, non standard. Peut-être en fait-il un usage entre guillemets, où « bon » signifie « approuvé par le code moral de ma société ». McNaughton reproche à cette solution de reposer sur la distinction entre significations descriptive et évaluative, contre laquelle il a argumenté plus tôt (1988 : 52-53). Selon lui, le fait que les propositions morales soient employées pour évaluer les actions n'implique pas qu'il y a deux actes illocutoires – description et évaluation – à distinguer et qu'elles relèvent du second. Il prétend que d'un seul acte locutoire il est possible à la fois d'évaluer et de décrire. En disant « Il y a un taureau dans le champ », à la fois l'émetteur décrit la situation et prévient le récepteur d'un danger. Mais cet argument est fallacieux. Il confond actes *illocutoires* – ce qu'on fait en disant quelque chose – et actes *perlocutoires* – ce qu'on essaie de faire en disant quelque chose, l'effet escompté. La distinction entre signification descriptive et signification évaluative relève de la première catégorie, les actes *perlocutoires* n'ayant rien à voir avec la signification à proprement parler (ils relèvent de la pragmatique plutôt que de la sémantique). Quand on dit « Il y a un taureau dans le champ », on décrit la situation (acte *illocutoire*) pour que le récepteur s'enfuit (acte *perlocutoire*).

Une autre raison pour laquelle McNaughton s'oppose à la distinction est que, selon lui, il existe des états mentaux mixtes, les croyances motivationnelles. S'il existe de tels états, pourquoi ne nous contenterions-nous pas d'un acte de langage mixte ? Mais comme nous l'avons vu, les croyances motivationnelles n'existent pas. Il faut concéder que si les opinions morales ne sont pas des croyances motivationnelles, l'évaluation et la description sont deux actes mentaux distincts et que cette distinction devrait se refléter dans le langage. Ce qu'elle fait si et seulement s'il existe deux significations, l'une descriptive, l'autre évaluative.

Jusqu'ici, notre réponse a quelque chose d'*ad hominem*. Un externaliste pourrait accepter la distinction entre significations descriptive et évaluative, mais nier que les énonciations morales sont des actes d'évaluations. Mais si cela devait constituer un argument, sa circularité serait par trop flagrante : on ne prouve pas la fausseté de la thèse internaliste en affirmant que les énonciations morales ne constituent pas des actes d'évaluation – *i.e.* en affirmant la thèse externaliste. N'oublions pas que la charge de la preuve repose sur les épaules de l'externaliste.

L'amoraliste dont les énoncés moraux ne sont pas évaluatifs ne pose donc pas de problème au cognitiviste. Si par contre ses énoncés moraux sont évaluatifs, alors nous avons affaire soit à un hypocrite – qui prétend désirer alors qu'il n'en est rien ; qui n'a pas d'opinion morale –, soit à une victime de la faiblesse de la volonté plutôt qu'à un amoraliste. Dans tous les cas, l'amoralisme ne permet pas d'invalider l'internalisme.

La même réponse, en termes de significations, conviendra également pour le cas du méchant. Il utilise « mal » dans un de ses sens descriptifs. Par exemple en retenant la signification descriptive et les conditions d'applications du terme adoptées par sa société, ou en imaginant ce à quoi il aboutirait s'il l'utilisait dans un sens évaluatif – probablement des conclusions utilitaristes, pour comprendre « bon » comme « maximisant l'utilité ». À côté de cela, il éprouve un désir de faire ce qui est mal, dans ce sens descriptif – et c'est un fait contingent. Bref, rien de bien embarrassant non plus pour l'internalisme.

Pourquoi il ne faut pas se contenter de l'internalisme atténué

Faut-il se satisfaire de la version de l'internalisme à laquelle nous avons abouti – l'internalisme atténué de Davidson ? Deux raisons, principalement, s'y opposent. Premièrement, les arguments par inférence à la meilleure explication ne sont pas tout à fait satisfaisants. Et c'est le seul moyen d'établir cette forme d'internalisme. L'attribution du statut de « meilleure explication » est toute relative. Ne serait-ce que temporellement : difficile de s'assurer que le camp adverse ne sera jamais capable d'en fournir une encore meilleure. Et puis, une inférence de ce type n'aura jamais la force d'un bon vieil argument déductif. Ensuite, l'internalisme de Davidson ne permet pas de défendre certaines formes de non-cognitivism contre la menace de la faiblesse de la volonté. Le prescriptivisme, pour ne citer que lui, requiert que les opinions morales soient accompagnées de motivations absolues, et non seulement *prima facie*. (Rien n'empêche les désirs de s'opposer au sein d'un même esprit. Par contre la souscription à deux prescriptions contraires est une faute logique, car les prescriptions sont les expressions de désirs *tout bien considéré*. En conséquence, les désirs

exprimés par les propositions morales ne peuvent être contrés par d'autres désirs.) Malheureusement, les explications de la faiblesse de la volonté fournies par Hare peuvent paraître inadéquates, parfois *ad hoc* (Hare 1999b ; Davidson 2008 : 45-46 ; Evans 1975 ; Lukes 1965 : 106-7). Il faut donc examiner les alternatives à l'internalisme atténué. A cette fin, la distinction suivante, entre trois formes d'internalisme, ne devrait pas être inutile :

- (Int1) *De facto*, les opinions morales sont intrinsèquement motivantes.
- (Int2) *De facto*, certaines opinions morales sont intrinsèquement motivantes, et ce sont ces opinions, et seulement ces opinions, qu'il est pertinent d'analyser (en méta-éthique).
- (Int3) Quoi que soient les opinions morales dans les faits, elles devraient être intrinsèquement motivantes.

Si l'on nous accorde le reste de notre argumentation – il est impossible de passer entre les cornes du trilemme (P1) et la théorie humienne de la motivation est vraie (P2) –, (Int1), (Int2) et (Int3) impliquent respectivement trois formes de non-cognitivismes :

- (NC1) *De facto*, les opinions morales sont des états mentaux non cognitifs.
- (NC2) *De facto*, certaines opinions morales sont des états mentaux non cognitifs, et ce sont ces opinions, et seulement ces opinions, qu'il est pertinent d'analyser.
- (NC3) Quoi que soient les opinions morales dans les faits, elles devraient être des états mentaux non cognitifs.

Clairement, la deuxième variante d'internalisme (et, donc, de non-cognitivismes) est plus faible que la première. Et la troisième, plus encore. Nous suggérons que si doute légitime il y a, c'est à propos de (Int1), puisqu'il pourrait ne pas parvenir à rendre compte de certains cas de faiblesse de la volonté. Le doute pourrait atteindre (Int2). A condition toutefois qu'on soutienne que soit toutes les opinions morales sont intrinsèquement motivantes, soit aucune ne l'est (*i.e.* de nier son premier conjoint). Ce qu'on aurait tort de faire. C'est d'ailleurs l'un des principaux reproches que l'on pourrait adresser à (Int1), ainsi qu'à la forme d'externalisme correspondante : il est très plausible que les opinions morales oscillent – d'un individu ou d'un instant à l'autre – entre des états motivationnels et des états inertes du point de vue de la motivation. En admettant que la force illocutoire des propositions morales fût à l'origine prescriptive, leur grammaire superficielle – la forme prédicative – est, traditionnellement, celle des propositions descriptives. Et c'est peut-être ce qu'exige le développement d'une culture morale, la reconnaissance de l'autorité des exigences morales,

etc. Ainsi, la souscription à une proposition morale pourrait, dans de nombreux cas, être vécue comme un acte cognitif. Va pour le premier conjoint de (Int2) ! Passons au second.

L'importance des attitudes morales est ce qui unit (Int2) et (Int3). Ce dialogue très réaliste devrait permettre aux plus récalcitrants d'en prendre conscience :

François : Se moquer des personnes âgées est mal.

David : (*l'air socratique*) Ah, oui ? Et pourquoi donc ?

François : Les personnes âgées sont des êtres humains. Et, donc, dignes de respect.

David : La notion de dignité humaine n'est qu'un produit de l'oppression des animaux non humains par les humains. (*David est antispéciste.*)

François : Alors disons que l'utilité totale s'en voit diminuée.

David : Explique-toi.

François : Eh bien, les personnes âgées ont des préférences, parmi lesquelles celle qu'on ne se moque pas d'elles.

David : Tu m'as convaincu : il ne faut pas se moquer des personnes âgées.

François : (*après le temps de réflexion qu'impose ce genre de situations*) Bon. McDo ?

David : Allons plutôt nous moquer de Bobby !

Bobby est un vieil homme. Que penser du projet de David ? Ne le trouvons-nous pas un peu décevant ? Ne préférons-nous pas que le changement d'opinions morales ait occasionné un changement d'attitudes, une motivation à agir différemment ? Et, si c'est le cas, ne préférons-nous pas que ce soit toujours le cas ? Qu'il soit impossible qu'il en aille autrement ? Si oui, nous devons préférer un langage moral prescriptif à un langage moral descriptif. Nous devons espérer que les jugements exprimés par ce langage soient des désirs et non de simples croyances. (Int2) affirme que c'est parfois le cas, ce que se garde de faire (Int3). Mais l'essentiel, c'est que ces deux formes d'internalisme reconnaissent l'importance des attitudes morales et l'avantage que constitueraient leur intégration dans la conception que nous nous faisons de nos opinions morales et la généralisation de cette intégration. Le langage n'est pas immuable. Il est affaire de conventions – qui sont enseignées –, fonction des pratiques linguistiques. Il serait non seulement *moralement* avantageux d'adopter des conventions linguistiques prescriptives quant à la morale – c'est presque une tautologie : les conclusions des discussions morales seraient nécessairement mises en pratique –, mais aussi d'un point de vue strictement prudentiel.

Ainsi, nous suggérons l'adoption de (Int3), qui ne contredit pas l'externalisme dans sa forme traditionnelle (« (Ext1) », pour faire court) sans se passer pour autant des prescriptions méthodologiques de (Int2). En cela, nous proposons non pas d'abandonner l'internalisme atténué de Davidson, mais d'accepter les deux versions – puisque la logique n'y voit pas d'objection.

4. Réalisme et non-cognitivism ne sont pas inconciliables (P4)

McNaughton, pour justifier la confrontation qu'il organise entre réalisme et non-cognitivism moraux, affirme que le cognitivism moral implique, et est impliqué par, le réalisme moral. Autrement dit, que les propriétés morales existent si, et seulement si, les opinions morales ne sont rien de plus que des croyances, ne requièrent pas l'existence de désirs. À l'appui de cette affirmation d'équivalence, il propose l'argument suivant (1988 : 7) – formalisé par nos soins :

(R1) Si x croit que Fa , x croit que Fa est vraie,

(R2) (Or) si Fa est vraie, alors il y a une propriété F telle que a possède F .

(C3) (Donc) si x croit que Fa , x croit qu'il y a une propriété F .

Lequel pose plusieurs problèmes. Nous en sélectionnons six.

Premièrement, il semble qu'accepter la première prémisse revienne à commettre une pétition de principe. Elle paraît difficilement contestable dans le cas des propositions descriptives, mais rien ne dit qu'il en aille de même des propositions morales. Sauf le principe de parcimonie, admettons. (Qui dit toutefois seulement qu'il faut, *ceteris paribus*, penser que c'est le cas.)

Deuxièmement, la seconde prémisse traduit la souscription à une conception particulière de la vérité : celle de la vérité comme correspondance à la réalité, selon laquelle une phrase est vraie si et seulement si elle correspond aux faits. Et cette conception ne s'applique pas à tous les usages du concept de vérité. Par exemple, les énoncés logiquement vrais comme « Soit les propriétés morales existent, soit elles n'existent pas » le sont indépendamment de la constitution du monde. Il n'y a pas de faits qui puissent les rendre vrais. Idem pour les propositions analytiques, telles que « Aucun célibataire n'est marié ». (Certes, on pourrait s'assurer empiriquement qu'aucun célibataire n'est marié, mais ce serait passer à côté du fait que la proposition est vraie indépendamment de ce test fastidieux.) Ces vérités, logiques ou

analytiques, relèvent d'un domaine épistémique qui rend le test empirique inadéquat. Les énoncés atomiques modaux, métaphysiques posent le même genre de problèmes. Par définition, leur vérité ne saurait être testée empiriquement.

La théorie de la vérité comme correspondance ne permet donc pas de rendre compte de certains usages du terme « vrai ». Pourquoi faudrait-il dès lors penser qu'elle rend correctement compte de la vérité morale ? Ranger la vérité morale dans la catégorie « vérités dont la théorie de la correspondance rend correctement compte » revient à commettre une pétition de principe lourde de conséquences.

En outre, il existe une autre théorie de la vérité : la théorie déflationniste (ou minimaliste) de la vérité, selon laquelle « P » est vraie si et seulement si P . Ses partisans affirment que rien de plus ne peut être dit à propos du concept de vérité. Cette théorie a l'avantage de rendre compte des usages énoncés ci-dessus du concept de vérité. (« $P \rightarrow P$ » est vraie si et seulement si $P \rightarrow P$.) Et de le faire sans suppositions métaphysiques (à plus forte raison dans le cas qui nous concerne, où les présupposés métaphysiques sont plus lourds que partout ailleurs, n'en déplaise à McNaughton), ce qui accroît son attrait sous l'angle de la parcimonie. Il faudra par contre, si on l'accepte, rejeter la seconde prémisse de notre argument : « Fa est vraie » implique que Fa , mais pas qu'il y a une propriété F telle que a possède F . McNaughton (1988 : 32) rejette cette théorie sous prétexte qu'elle interdirait à la vérité de trancher les débats moraux : si dire que P est vraie revient à affirmer P , alors il est impossible d'établir de manière objective ce qui est moralement bon/mauvais. Mais attribuer ce rôle, et une telle responsabilité, à la notion de vérité revient à commettre – décidément ! – une pétition de principe, d'autant plus étonnante que McNaughton admet plus loin qu'il est raisonnable de penser que le non-cognitivisme implique l'utilitarisme, tout en délestant le concept de vérité de cette lourde tâche. Gageons qu'il nous concède que l'utilitarisme est, autant que n'importe quelle autre, une théorie morale objective.

Troisièmement, il semble que, même dans la conception de la vérité comme correspondance à la réalité, un cognitiviste pourrait tout à fait accepter (R2) pour les propositions morales. Les non-cognitivistes s'opposent aux cognitivistes non pas sur le fait que les propositions morales expriment des croyances, mais sur le fait qu'elles *n'*expriment *que* des croyances. Ils soutiennent, avec leurs opposants, que les propositions évaluatives ont une signification descriptive. Mais ajoutent qu'elles ont également une signification évaluative. Or si une proposition a une signification descriptive, elle possède par là même une valeur de vérité, et des conditions de vérité (*i.e.* des conditions qui la rendent vraies au sens de la théorie de la correspondance). L'argument de McNaughton ne peut être opposé qu'à celui

qui soutient que les propositions morales *n'*expriment *que* des désirs. Ce que (plus) personne ne soutient. Les réductionnistes également peuvent accepter (R2). Avec les non-cognitivistes, ils accepteront la conclusion (C3) avec le bémol que la propriété *F* en question n'est pas une propriété morale irréductible, mais une (conjonction de) propriété(s) descriptive(s) des plus banales.

Quatrièmement, même en oubliant les trois points précédents, l'argument n'établit pas le réalisme moral. La théorie de l'erreur, par exemple, souscrit à sa conclusion tout en rejetant le réalisme : affirme que notre croyance en l'existence de propriétés morales est systématiquement fautive. Oui, nous croyons en l'existence de propriétés morales irréductibles, et nos jugements moraux sont l'expression de ces croyances, mais, en cela nous avons toujours tort : les propriétés morales n'existent pas.

Cinquièmement, l'argument est invalide, du moins dans cette formulation. Ce que met en évidence sa formalisation (« *x* croit que *P* ; or $P \rightarrow Q$; donc *x* croit que *Q* »). La conjonction des prémisses sera vraie et la conclusion fautive si *x* n'adhère pas à la seconde prémisse. (Et possiblement même s'il y adhère, d'ailleurs : la seule conclusion logiquement autorisée est que s'il adhère à la seconde prémisse, il devrait adhérer à la conclusion – une prescription épistémique, l'obéissance à laquelle n'a rien d'une nécessité logique.)

Problème majeur, enfin : l'argument ne permettrait, si l'on choisissait de mettre de côté ces quelques difficultés, d'établir que l'implication du réalisme par le cognitivisme et non la relation inverse qui est nécessaire à l'argument antiréaliste que présente McNaughton. Cette relation – l'implication du cognitivisme par le réalisme – est également fautive. Mais pour une raison différente et plus simple. Même si des propriétés morales irréductibles existaient, pourquoi nos opinions morales devraient-elles nécessairement – logiquement – se limiter à des croyances ? On pourrait argumenter que si les propriétés morales existaient, il serait plus adapté de le croire plutôt que de cultiver des désirs indépendants de leur existence, mais c'est nettement au-dessous de ce qui est requis par le cognitivisme (*i.e.* que, *de facto*, nos opinions morales sont des croyances). En outre, pourquoi faudrait-il être convaincu ? Celui dont la pratique morale est non cognitive (celui dont les opinions morales sont des désirs), de même que l'amoral, pourrait répondre au réaliste cognitive « Oui, les propriétés morales existent, et alors ? » et se comporter en conséquence sans commettre la moindre infraction aux règles ni de la logique, ni de la rationalité, ni même de la morale ! (Mais c'est la logique qui importe ici.) Croire en l'existence des propriétés morales ne revient pas à exprimer des croyances par ses jugements moraux.

Le non-cognitivism, même dans sa forme traditionnelle (NC1), n'implique donc pas l'anti-réalisme. Le contraire eut toutefois été étonnant : comment une thèse sur la nature de nos opinions pourrait-elle impliquer quoi que ce soit quant à la constitution du monde ? On peut comprendre – sans la partager – la croyance que le réalisme implique le cognitivism (par exemple, si on l'explique par la confusion entre existence d'une croyance et légitimité possible d'une croyance : le réalisme implique la possible légitimité des croyances morales, pas leur existence). La croyance inverse dépasse par contre notre entendement.

Conclusion

Nous avons défendu les trois premières prémisses de l'argument antiréaliste : la conjonction de la théorie humienne de la motivation, de l'internalisme moral (atténué) et du cognitivism moral est incohérente ; la théorie de la motivation est vraie ; et l'internalisme moral (atténué) est vrai. Nous avons par contre établi la fausseté de notre quatrième prémisse. L'argument que présente McNaughton contre le réalisme n'est donc pas solide. Mais ce n'est pas pour la raison présentée par l'auteur – le rejet de la théorie humienne de la motivation –, mais en vertu de la fausseté de l'équivalence qu'il admet entre cognitivism et réalisme.

Quoi qu'il en soit, les trois premières prémisses nous permettent de rejeter le cognitivism. Et nous avons argumenté que la question épistémologique était plus pertinente que la question ontologique : si les opinions morales sont des états non cognitifs, alors, que des propriétés morales et des croyances en leur existence existent ou non, elles ne détermineront en aucun cas les opinions morales. Et pour cause : les propriétés déterminent quels états cognitifs sont adéquats, mais jamais rien quant aux états non cognitifs.

C'est l'internalisme atténué qui nous a permis de rejeter le cognitivism. Toutefois, et en guise de conclusion, nous soutiendrons que (Int3) est philosophiquement plus intéressant. En effet, il serait difficile de tirer des conclusions incontestables en éthique normative de l'adoption de l'internalisme atténué. Comme le souligne McNaughton, on peut considérer que le point de vue moral *est* celui du spectateur idéal, mais on peut aussi penser le contraire, puisque dans ce cas, l'identification des points de vue est un principe substantiel et non formel. Par contre (Int3) a des implications non négligeables : si, idéalement, les opinions morales devraient impliquer une motivation absolue, alors elles devraient exprimer des désirs tout bien considéré.

Admettons, pour les besoins de l'argument (et plus si affinité), ces deux thèses :

(Univ) Les propositions morales sont universalisables.

(Util) La conjonction de (Univ) et du prescriptivisme implique une forme d'utilitarisme.

Le prescriptivisme est la thèse selon laquelle les propositions morales expriment des désirs tout bien considéré. Par définition, donc, dire que les propositions morales devraient exprimer des désirs tout bien considéré, c'est dire que le prescriptivisme devrait être vrai – que nous devrions adopter un langage moral prescriptif. Puisque nous avons admis (Univ) et (Util), il faudra conclure que l'utilitarisme devrait être vrai. *I.e.* que nous devrions devoir maximiser l'utilité. Mais peut-on affirmer cela et nier que nous devons maximiser l'utilité ? Il semble que non. On pourrait arguer que le monde dans lequel nous vivons n'est pas idéal et que donc l'utilitarisme ne doit pas s'y appliquer. Mais tout ce que demande (Int3), c'est que le *langage moral* soit idéal. Le monde peut bien rester tel qu'il est pour le reste (en ce qui concerne (Int3)). On est en droit de se demander comment une simple différence de langage entre deux mondes possibles pourrait transformer de mauvaises actions dans l'un en bonnes actions dans l'autre.

Bien sûr la question reste ouverte, mais il est au moins plausible que, dans sa défense de l'utilitarisme, Richard Hare, qui avait délaissé (Int1) à la faveur de (Int2) en développant sa théorie des deux niveaux – les énoncés moraux comme descriptions au niveau intuitif, comme prescription au niveau critique –, aurait peut-être eu avantage à lui préférer (Int3), et donc un prescriptivisme affaibli en conséquence. Cette recommandation arrive malheureusement un peu tard, qui ne lui est désormais plus d'aucune utilité. Paix à son âme.

RÉFÉRENCES

- DAVIDSON, Donald (2008). "Comment la Faiblesse de la Volonté est-elle possible ?", in *Actions et Evènements*, PUF.
- EVANS, Donald (1975). "Moral Weakness", *Philosophy*, 50 : 295-310.
- FURRER, D.P. (2009). "Travail Ecrit", non publié.
- GARDINER, P.L. (1964). "Pain and Evil, II", *Proceedings of the Aristotelian Society*, 38 : 107-124.
- HARE, R.M. (1999a). "Internalism and Externalism in Ethics", in *Objective Prescriptions and Other Essays*. Oxford University press.
- (1999b). "Weakness of the Will", in *Objective Prescriptions and Other Essays*. Oxford University press.
- LUKES, Steven (1965). "Moral Weakness", *The Philosophical Quarterly*, 15 : 104-114.
- MCNAUGHTON, David (1988). *Moral Vision*. Blackwell.
- SMITH, Michael (1987). "The Humean theory of motivation", *Mind*, 96 : 36-61.